SOCIAL SCIENCE REVIEW ARCHIVES
https://policyjournalofms.com
ISSN Online: 3006-4708
ISSN Print: 3006-4694

Global E-Commerce Dynamics: Analyzing Cross-Border Consumer Behavior and Insights from Electronic Sales Data.

Dr. Malik Muhammad Ali[1], Muhammad Taqi[2], Prof. Jia Haitao[3], Farhad Javed[4]

[1] Assistant Professor. TIMES Institute Multan, Post-Doctoral Fellow Jinan University.
[2] Lecturer. TIMES Institute Multan, Email: muhammadtaqi@t.edu.pk ali_ir91152@yahoo.com
[3] Professor. Jainan University China
[4] Lecturer. TIMES Institute Multan, Pakistan

DOI: https://doi.org/10.70670/sra.v3i1.500

Abstract
This study explores the key factors influencing e-commerce sales transactions by analyzing electronic sales data from September 2023 to September 2024. Using regression analysis, the findings reveal that unit price and quantity purchased strongly predict total sales, with unit price having the highest predictive power. A positive correlation between quantity and total sales highlights the impact of bulk purchasing, while add-on totals show minimal influence, suggesting opportunities for improved marketing strategies for complementary products. Descriptive statistics indicate significant variability in customer spending, with a right-skewed distribution showing that most customers spend less, while a few contribute disproportionately to total revenue. ANOVA results confirm the significance of differences between stores, reinforcing the importance of targeted pricing and inventory strategies. However, limitations such as unaccounted variance (12.4%), static data, and the absence of qualitative factors like customer satisfaction suggest avenues for further research. Future studies should incorporate seasonal trends, customer segmentation, and machine learning techniques to enhance predictive accuracy and strategic decision-making in e-commerce.

Key Words: E-commerce Sales, Consumer Behavior, Marketing Strategy, Digital Commerce, Price Elasticity, Customer Segmentation

Introduction
Customer purchase behavior analysis is crucial to organizations that seek to adopt better business practices and improve customer fulfillment. In the period specified above, the Electronics industry has grown exponentially, and customer demands have forced innovations and competition in the market. Purchases also present a rich delimited field of how demographics, product categories, and buying behavior patterns influence many decisions. This project employs a wide database from September 2023 to September 2024 to analyze relevant factors affecting the sales transactions of an electronics firm. All these questions are important in tackling present-day research queries, which are in tandem with contemporary issues in retail. Therefore, this work intended to utilize the dataset to reveal insights, discover possibilities of delays, and make recommendations to increase sales effectiveness(Biswas et al., 2024). Such demographic variables like age, gender, and membership to loyalty will contribute towards enhancing the understanding of how these affect the total purchase value. Earlier studies proved that personality factors rather play crucial roles in influencing purchase control and consumers' allegiance. By knowing the patterns of these demographics with spending, appropriate customer groups can be segmented, and proper marketing steps can be taken. Of particular interest is the ability of loyalty

programs to encourage consumers to make repeated purchases, and evaluating this mechanism over time can be done with the help of this dataset. Some of the research questions include: Is there a significant difference between loyalty members and non-members? Sex, as many researchers may not consider it, can also facilitate understanding of the spending disparities and maybe help in giving recommendations regarding products. Exploring these issues provides more insights into the nature of demographic attributes influencing financial performance of retailers(Black, 2024). According to product rating, two other research questions are relevant to customer satisfaction. Compared to other response measures, rating entails the identification of the quality, nature, and intensity of certain products or services and customers' experiences, which is critical. Studies have found out that customer satisfaction has a close relationship with repeat purchase and brand loyalty. This dataset enables us to compare ratings of products in different such as Smartphone, laptops, and tablet. Just as in any other retail environment, trends in satisfaction can be used to pinpoint areas in which certain product lines need to be enhanced. In addition, analyzing ratings could let companies understand which attributes of the services and products customers cherish most, thus aiding in product design. The interaction of the price with satisfaction is another issue of interest, as consumers think that higher prices must mean better quality products. Thus this study has a usefulness in studying these correlations to shed light on consumer's preference and expectation levels(Faiza & Taher, 2024). Another aspect for discussion in this report is the shipping possibilities as well as their relationship to the completion rates of orders. Specific examples of shipping options are "Standard," "Overnight," and "Express," which have an impact on customers' decisions and their perceptions. Studious show that effectiveness in shipment can greatly boost customer satisfaction and reduce cancellations. Hence, the dataset extraction was useful to determine the differences in shipping preferences of customers and their relation to the orders. Comparing cancellations trends by types of shipping can reveal originality, overarching problems, and potential opportunities for refining the process. Furthermore, many of these costs with regards to order size and value can be used as a way of understanding customers. Such information should help businesses to develop adequate shipping policies to best meet their objectives in a competitive market. This study also provides some insight into the supply chain management practice of e-commerce companies(Hashemian et al., 2024). Payment options are also an important aspect of customer experience and tendencies affecting the probability of customers' purchase decisions. This layer contains a variety of payment types, including cold, credit, and electronic, including the PayPal bill; knowledge of payment preferences across demographics can aid in improving the levels of convenience that are offered. Tendencies in payments can also reveal new technologies; for example, the number of people using digital wallets has risen sharply over the last few years. The evaluation of payment options can also help discover whether some of them are characterized by a high rate of cancellations. Findings from this analysis can be used to inform how businesses approach the incentives, such as giving a discount for one method of payment over the other. Furthermore, by comparing current information on payments with information collected in the past, new trends in terms of consumer trust and technological trends can also be identified. Hereby, the mentioned aspects help to complement the understanding of transaction dynamics resulting from the given report (Khuan et al., 2024). The additional purchase factors include accessories and warranties that make it difficult to determine the true nature of purchase behavior. Hoods are usually part of schemes as value additions and the trends in realized sales of hoods indicate consumer appetite for such products. From the data analysis, it could be established on the cross-tabulation of the products and the Add-ons how often add-ons are bought along product categories. This information is very valuable for marking the cross-selling and upselling techniques used in the retail facilities. Additional parameters that can be considered in the framework of the identified metrics include add-on prices and their share in the total purchase prices. For instance, customers with heavy capital invested in warranties may have developed the need for long-lasting products. This information will allow consumers to know which groups are more likely to purchase add-ons, thus creating better customer leads. Moreover, an overall and detailed analysis of add-ons through time can provide information about their periodicity and promotional success. When these factors are included, the business shall have improved its product bundling options and increased the gross revenue collected.(Madyatmadja et al., 2024). Another important aspect examined in the present research is temporal trends in sales transactions. Analyzing purchase dates can tell us when the business is busiest when it had major sales and promotions, and when customers' needs change. Reviewing sales data for the year may reveal more periodic sales times, such as during holiday seasons or back-to-school periods. These are trends which are crucial for effective planning of inventory and promotional crusades. The dataset also gives an insight on how factors outside the immediate control of the business, like economic factors affect the rates of sales. It is most useful to divide sales data by months so that a business can understand how certain activities and promotions affect its results. The dynamism in customer behavior by time can help the businesses to develop time-sensitive products, discounts and promotions for their products. Hence temporal trends can also be used to identify specific periods of underperformance that should be corrected for. This analysis not only refines the company's ability to sales forecast, but also the long-term strategic planning(Raji et al., 2024). Also as we explore in this report another focus is the understanding of consumer's price elasticity. The information of price variable of the data set is highly elaborated to understand how the customers are sensitive to prices. Knowledge of price elasticity can help firms set right prices in relation to their rivals as well as develop effective discount policies. It gives business standings an insight into the right quantity of purchase in relation to varying unit prices that yields maximum returns. The dataset also allows a comparison of average prices by a group of products, which could help to indicate gaps in the value chain. Price sensitivity on demographic factors can therefore be of more help in evaluating customers' tendencies. For instance, price sensitivity tends to vary over the type of customer, and therefore one may find that customers from the younger generation are more sensitive to price changes than individuals of older generations. Furthermore, it is equally possible to analyze the price change over some period and identify changes in consumption pattern or market factors. It is with this understanding that this analysis seeks to advance knowledge in the field of price and buying behavior relationship(Sunarya et al., 2024). The information available on e-commerce is rich with data and its analysis provides the ability to learn more about customers, products, and operations. Using a large database of electronic sales from September 2023 to September 2024, this research aims to examine the impact of numerous factors that include both quantitative details and qualitative characteristics of the customer, including membership, type of product, and shipping facilities requested on the performance of the value driver sales and customer satisfaction score. To this end, the study aims to answer three research questions: What is the effect of demographic factors and loyalty programs on total sales? How do product types and add-on contributions affect the transaction amount? And, how does segmentation by shipping method affect order conversion rates? The findings of this study provide a richer understanding of the subject of e-commerce analytics but are also useful to inform advanced customer engagement and sales growth. As such, employing various statistical methods and an empirical research framework, this paper ensures a major contribution to both theoretical knowledge of e-commerce and the delivery of efficient tools for practicing managers in today's rapidly changing world.

Research Questions

1. How do demographic factors (age, gender, and loyalty membership) influence the total purchase value?
2. What is the relationship between product type and customer satisfaction (product rating)?
3. Does the choice of shipping type impact the likelihood of order cancellation?

Literature Review

E-commerce has come of age as a disruptive technology in the retail industry since it is altering the manner in which retailers relate to their consumers. Regarding the e-commerce customers' behavior, Sunarya et al. (2024) showed that machine learning approach is suitable, especially when the random forest algorithms are preferable to the logistic regression on the aspect of the purchase tendency. Their study also shows the need for utilising advanced analytics in business models to unlock consumer engagement insights from digital domains. Tarnanidis (2024) stressed that mobile marketing provides great impact to extra-push decisions because of targeted and relevant ads of products being Pope (2016), advertisement campaigns which are focused spatially as well as in relation to customer interests significantly increase purchasing decisions. Mobile marketing strategies are those that allow communicating, at the right time, relevant messages to consumers, an element essential in today's competitive e-commerce market. In combination, these papers offer evidence of how big data and marketing communication can enhance engagement and sales for a firm in the modern world. There has been a lot of research done on how demographic factors affect consumer purchasing behavior for better customer classifications and market targeting. Yang et al. (2023), in a study on factors that motivate e-shopping behavior using a multi-analytic method, gathered evidence that confirmed convenience, trust, and product differentiation as significantly influential on the e-shopping decision. Their studies are in tune with other studies' suggestions that there is a need to have measures that foster trust, for instance, safe ways of paying and clear refund policies. Thus, supported by the findings of Khuan et al. (2024), it is possible to conclude that consumer perceptions of product quality and their ease of buying certainly remain vital parameters for decision-making within start-up environments. Indeed, the multiple-mediated model indicates that increasing customer value perception and reducing decision-making complexity is paramount to customer satisfaction and loyalty. These observations raise the issue of demographic-centered initiatives as the key to developing more targeted and effective ways of shopping. Technology serves as one of the key factors that determine customer experiences and business throughout e-commerce operations. Zhang (2024), the author probed into exploring the positive changes of the effect of AI in retailing, particularly on the customer side, which encompasses AI-coupled customer recommendations and robotic customer services. Raji et al. (2024) also presented tendencies of AI personalization in relation to the role of adaptive technologies for new and returning clients. Black (2024) reviewed the possible analytical techniques based on clickstream data and sentiment analysis for factors in the conversion of purchasers online, showing how AI tools can bring the complexity of clients' behaviors. Altogether, these works highlight the inability of e-commerce to improve its practices and develop closer connections with clients, in part from the lack of technology support. It is, therefore, safe to conclude that AI and machine learning are the key drivers of the current e-commerce strategy and operation. Customer satisfaction and quality retention continue as some of the key goals of e-commerce since it is a common business goal that any e-commerce business aims at retaining current customers whilst attracting new ones. Like Biswas and colleagues, we also identified a shift towards omnichannel behavior in in-store and online shopping channels, Biswas et al. (2024). From the study, it is clear that omnichannel shopping helps develop a consistency of shopping that acts as a customer satisfactoriness factor. In addition, Hashemian et al. (2024) addressed customer segments and A/B tests in e-book platforms and underlined the necessity of studying customers to improve subscription sales. These findings support the earlier works of Khuan et al. (2024) on the importance of quality perception and promotional strategies in building customer loyalty. In identifying the antecedents of satisfaction and consequent retention, these studies add more depth to the literature on future sustainable customers for e-commerce Log inactivity& Knowledge & Gender & Relevance. Pricing and promotion strategies are key elements of the buying processes and as such, their understanding is central to explain, and possibly shape, consumer behavior. The analysis study of the marketing logistics and consumer behavior in the context of the Indian e-commerce industry was done by Tripathi et al., where they elaborated on how competitive pricing attracts customer interest. Moreover, Sunarya et al., (2024) showed that consumers are okay with the dynamic pricing strategies once supplemented with data-driven promotions. Tarnanidis (2024) also identified that m-Marketing mobile campaigns that promote exclusive offers have a higher CTR and conversion rate. All these studies together also underscore the significance of pricing decisions for customers' perceptions and overall sales. It becomes easier for a business to develop an effective pricing strategy when they know how customers are likely to respond to specific types of promotions that they have instilled in the market. The involvement of product characteristics and attachments in the purchase decision has also received consideration in e-commerce studies. Thus, in,Madyatmadja et al. pointed out that stimulation of social media purchase features as product appearance and users' feedback elicited higher perception of purchasing students. Hashemian et al. (2024) observe that adding warranties and accessories is deemed to be favorable to the overall quality of a product. Similar conclusion can be made with the findings of Khuan et al., (2024), who also noted that purchase add-ons enhance overall purchase values. The rational behind this is by emphasizing on product improvement and related services, organizations are able to develop customized product solutions that address the needs of different customers in value net. This strategy not only enhances the sales-revenue figure, but also enhances customer value as well. Seasonality and promotional timing are some of the temporal factors which are important in achieving an understanding of e-commerce sales trends. Yang et al. (2023) noted that more traffic is generated in the respective sites mainly due to increase in traffic during festive seasons and sale periods like black Friday. They also confirm the conclusions of Tripathi et al. (2024), who established that Indian consumers are exceptionally sensitive to festival discount and flash sale signals. The general implication, therefore, is that promotions must be integrated into a firm-specific promotional calendar to capture customers during these high demand periods. Other aspects that Tarnanidis (2024) also focused on were real-time notifications as an instrumental factor that helps in making impulse-specific buying during promotions. Collectively, these papers offer a guide on how maximise the use of temporal knowledge within the marketing communications mix so as to plan and execute successful campaigns, and, in turn, drive increased demand in the all-important sales promotion periods. Transportation and payment factors define the possibilities to make complete transactions and increase customer satisfaction in the context of e-commerce. Sunarya et al. (2024) also pointed out that customers are likely to be retained by speedy and reliable shipping and pointed out that delayed orders lead to order cancellation. Concerning payment methods, Payment preferences as described by Raji et al (2024) shows the general population is gradually shifting towards the use of mobile wallets and Paypal. These preferences are in line with Black's (2024) which showed that extended payment modes improve trust and transaction effectiveness. In addition, Khuan et al. (2024) suggested that efficiency in payments affects positively customer satisfaction most especially in emerging regions. As discussed earlier, such concerns have to do with logistics and the ability or inability to satisfy the diverse payment wishes of customers can open a business up to creation of more dependable organizational structures. Cultural and technological factors are the two most significant sources of influence in consumer behavior in e-commerce. Biswas et al. (2024) emphasized cultural factors in relation to these preferences and mentioned that such subjects are valuable for analyzing new markets, common among which is cultural diversity. Madyatmadja, Kuppens, and Leichty (2024) found that the students are more likely to engage with social media selling features and peer suggestions. From these analyses, implications are raised pointing to the need to solve for culture and generation in the creation of e-commerce initiatives. The need was underlined by Raji et al. (2024) for another market-specific trend: artificial intelligence-powered personalization. Collectively, these works emphasize the need to be harmonious with e-commerce strategies and its target consumer base and their expectation levels. Last but not the least; the future research needs to take into consideration all those limitations and challenges which are emerging with the growth of this concept of e-commerce. Zhang (2024) stressed the risks of ethical and data protectionists in artificial intelligence where business has toconsider between innovation and acceptance. Hashemian et al. discussed this to the current challenges of segmentation and A/B testing by pointing out that they need to be further updated from one period to another due to shifting consumer behaviors. Concluding its findings, Tarnanidis (2024) equally pointed out the problem of maintaining extensive monitoring of mobile marketing to foster enduring participation. Its nature and development, as well as the trends in technology, necessitate the research to be adaptive and to look ahead. Therefore, it is possible for future research to build on the existing literatures by elaborating how consumers interact with marketing messages in the current world.

Hypotheses

Hypothesis 1: Impact of Customer Demographics on Sales Performance

H_0 (Null Hypothesis): Customer demographics (age and gender) do not significantly influence total sales performance (Total Price).

H_1 (Alternative Hypothesis): Customer demographics (age and gender) significantly influence total sales performance (Total Price).

Hypothesis 2: Influence of Product Types and Add-on Purchases on Transaction Value

H_0 (Null Hypothesis): Product types and add-on purchases do not significantly affect the overall transaction value (Total Price).

H_1 (Alternative Hypothesis): Product types and add-on purchases significantly affect the overall transaction value (Total Price).

Hypothesis 3: Relationship Between Shipping Type and Order Completion Rates

H_0 (Null Hypothesis): Shipping type is not significantly associated with order completion rates.

H_1 (Alternative Hypothesis): Shipping type is significantly associated with order completion rates.

Methodology

In this study, the strategy of data analysis is structured as analyzing the sales transaction data of an electronics company for one fiscal year starting from September 2023 to September 2024. It is about recognizing such essential variables affecting sales growth, customer satisfaction, and purchasing behavior in order to arrive at useful findings for enhancing e-commerce goals. It covers the data definition, the variables, and the methods of data analysis.

Data Description

The data held in the dataset covers a record of the sales transactions on various electronic products, a record of the customers' personal characteristic features, their buying habits and behaviors, and the product characteristics. The data used in this paper is obtained from the Kaggle platform and is named Customer Purchase Behavior - Electronic Sales Data (Sep2023-Sep2024). Among others, it contains information on 15 main attributes, including Customer ID, Age, Gender, Product ID, Payment Type, Date of Purchase, and Shipping Type. Speaking of missing values, there are none for some important independent variables such as customer ratings.

Variables

A. Dependent Variable:

Total Price (Numeric): The total dollar amount of a sales transaction that is used to benchmark sales productivity.

B. Independent Variables:

1. **Customer Demographics:**

Age (Numeric): Age of the customer is thought to affect the consumer buying power and their consumption behaviors.

Gender (Categorical): Male or Female, because the purchasing behavior will be compared between genders.

2. **Customer Membership and Engagement:**

Loyalty Member (Yes/No): Signals the customer to be a member or not of the loyalty program which affect its repeat purchase and expenditure.
Customer Ratings (1-5 stars): Bins evidence of product satisfaction which determines future purchase behaviors.

3. **Product and Pricing Variables:**

Product Type (Categorical): Smartphone, Laptop or Tablet and how it impacts the average order value.
Unit Price (Numeric): The amount of product that can be purchased for the available price level, helping explain total price differences.
Quantity (Numeric): Quantity of a particular item purchased at a time and is directly proportional to the total amount of money that has been agreed to be paid for those products.

4. **Transaction Characteristics:**

Payment Method (Categorical): Payment type that was applied (Cash, Credit card, PayPal etc.) as it enables the assessment of attitudes in relation to payment services.
Shipping Type (Categorical): Shipping type (Standard, Overnight, Express) which was believed to affect customer satisfaction with the order and order completion rates.

5. **Add-ons and Enhancements:**

Add-on Total (Numeric): The summed up employed amount or the monetary value of extra features which influence upselling, for examples warranties or accessories.

Results Analysis

Demographic Analysis

Table 1: Demographic Analysis

Gender	Frequency	Percent	Valid Percent	Cumulative Percent
Female	9835	49.2	49.2	49.2
Male	10164	50.8	50.8	100.0
Total	20000	100.0	100.0	

According to the customers' distribution by gender, it has been observed that male customers are slightly more in number than the female ones. For instance, out of 2000C= 20,000 customers, 1000M = 10, 164 customers are male representing 50.8% of the total customers while 2000F= 9,835 customers are female representing 49.2 % of the total customers. This means that unlike majority of companies the current Firm targets both the male and female in almost equal proportions; hence its marketing and products offering are likely to be suitable to the broad market. The valid percent for males and females matches overall valid percent because there are no mis sing or omitted data for this factor. Moreover, the cumulative percent also reaffirms that there is a representation of female customers as 49.2% and male's customers 50.8% combined give 100%. This 50:50 gender distribution might be useful to study how different genders approach the product and which specific promotions might be aimed at each sex. At the same time, the proportionate representation guarantees that the findings made with reference to gender comparisons are valid and representative of the entire customer population.

Descriptive Statistics

Table 2: Descriptive Statistics

Total Price		Unit Price	
Mean	3180.133418	Mean	578.6319
Standard Error	17.99571679	Standard Error	2.208111
Median	2534.49	Median	463.96

Mode	3932.05	Mode	786.41
Standard Deviation	2544.978675	Standard Deviation	312.2741
Sample Variance	6476916.456	Sample Variance	97515.1
Kurtosis	0.288780655	Kurtosis	-0.70823
Skewness	0.903509728	Skewness	-0.02703
Range	11376.05	Range	1118.93
Minimum	20.75	Minimum	20.75
Maximum	11396.8	Maximum	1139.68
Sum	63602668.37	Sum	11572637
Count	20000	Count	20000
Confidence Level (95.0%)	35.27309156	Confidence Level (95.0%)	4.32808

Quantity		Add-on Total	
Mean	5.48555	Mean	62.24485
Standard Error	0.0203	Standard Error	0.410535
Median	5	Median	51.7
Mode	3	Mode	0
Standard Deviation	2.870854	Standard Deviation	58.05843
Sample Variance	8.241803	Sample Variance	3370.781
Kurtosis	-1.22651	Kurtosis	0.392988
Skewness	0.001913	Skewness	0.936524
Range	9	Range	292.77
Minimum	1	Minimum	0
Maximum	10	Maximum	292.77
Sum	109711	Sum	1244897
Count	20000	Count	20000
Confidence Level (95.0%)	0.03979	Confidence Level (95.0%)	0.804683

Looking at the descriptive statistics for Total Price, the mean of transaction value is 3,180.13 with standard deviation of 2,544.98, showing that the transaction amounts vary a lot. Averagely, each transaction involves 2,534.49 which is higher than the median showing that this distribution has a positive skewness of 0.90. This is an indication that many of the transactions occupy the small and lower end of the scale, but the presence of several outliers shifting the mean in the process. The transaction values range starts from 20.75 and goes up to 11,396.80. By using kurtosis, the test result shows that the distribution is slightly more peaked than normal distribution and the kurtosis value obtained is 0.29. Of most importance is the 95% confidence interval for the mean, which comes out to be 3,180.13 ± 35.27 thus giving a true value of the population mean. As of Unit Price, the mean located at 578 .63 which has a lower value of standard deviation at 312 .27 as compared to Total Price. The median value is 463,96 which is again lower than the mean value, additionally there is negative skewness equal to- 0.03. The variety from the least value of 20.75 to the greatest value of 1,139.68 substantiates this stew. All the coefficients are below zero, specifically, skewness is -0.07, kurtosis is -0.71 meaning it is less peaked than normal. Taking the confidence interval into consideration, the mean of Unit Price is confidence level 95%, indicating high accuracy of estimating the average unit price 4.33 at 95%CI, 578.63 ± 4.33.

For the Quantity the arithmetic mean quantity purchased per transaction = 5.49 with a standard deviation of 2.87 the variability is moderate. Median and mode are 5 and 3 respectively; thus it shows that a smaller amount is more frequent than a larger one. The range of transactions is limited to 1 to 10 transactions per unit. It shows that skewness is equal to 0.002 – which means it is almost normally distributed the kurtosis is -1.23 which pointing towards flat distribution. The 95% confidence interval for the mean quantity is 5.49 ± 0.04 which confirms a good precision of the mean estimate. About Add-on Total, the mean of the additional bought product is 62.24 with a SD of 58.06 which indicates much variability in add-on sale. The arithmetic mean add on total is 35.95, a good number of transactions had no add-ons with a mode of 0 add-on total. This scope makes sense and it ranges from 0 to 292.77 which means that add-ons can have different values. The positive value 0.94 also signifies that majority of add-on totals are low, with a few a few extremely high values. The kurtosis is 0.39 indicating that the distribution of the returns must be slightly peaked. Which is 95% C.I., the Mean add-on total is equal to 62.24 ± 0.80 therefore the confidence interval will provide rather precise estimation.

Correlation Matrix

Table 3: Correlation Analysis

	Total Price	Unit Price	Quantity	Add-on Total
Total Price	1			
Unit Price	0.673951	1		
Quantity	0.653872	0.006714698	1	
Add-on Total	0.083924	0.125188978	0.003419	1

The correlation matrix reveals the interaction between total price or the unit price, quantity and the add on total. The highest coefficient equals 0.673, which shows that total price is moderately positively related with the unit price. From this it can be inferred that as the unit price of the items rises, the total price also rises in tandem. This relationship is in proportion with the fact that many expensive products are responsible for high total sales figures. Likewise, authors found that total price and quantity had a correlation of 0.654, meaning that the total price increases with the increase in the quantity purchased. This relationship brings out the effect of making bulk purchases or multiple item transaction in relation to total sales. There is a very low relationship between unit price and quantity and this is shown by a correlation coefficient of 0.007 suggesting that there is actually no relationship between the two. This means that fluctuations in the unit price in fact do not have any major effect on the quantity of the items bought, meaning that the customer quantity purchase behavior is only weakly associated with the price per item. This could bring to spotlight other factors like the necessity feature of the product, the availability or price dictatorship by the customers' budget. Notably, the add on total has low correlations with all variables showing an inverted U shape. Add-on total and total price have a very small and positive coefficient of determination of 0.084. This means that case accessories or further acquisition significantly do not affect the total price. Likewise, there is low positive relation between add-on total and unit price at a coefficient of 0.125 which is still very weak. This may also mean that all add-ons are not necessarily associated with the higher-priced category because of the scarcity of bundling or upselling techniques. Indeed, the level of association between add-on total and quantity is the lowest, 0.003, thus suggesting that there is no significant relationship between add-on total and add-on quantity. This goes further in supporting the neglect of add-ons on overall sales targets or rather key performance indicators and the strategic improvements that are most important are highlighted in the matrix. Although, add-on total is almost equally dominated by both add-on total and total price, the proved weak correlation within add-on total and other aspects may indicate that more attention should be paid to the utilization of add-ons in sales promotion. It is possible for firms to employ focused selling tactics, for instance, package additional accessories when one of the highly demanded products is sold or allowing a buyer to get other related product at a discount price as a way of enhancing its influence in total sales. Also, the poor relationship between the unit price and the quantity further imply that there is degree of variation in the price elasticity of demand You need to carry out further analysis to understand the differences and hence set right price for each segment of the market. As shown by the correlation matrix, total price has a close positive relationship with unit price and to a lesser extent with quantity, while add-ons terms have little effect. It offers important advice on the price and sales directions while pointing at some opportunities, like add-ons, which could be further developed to have a more positive impact on the company's revenue.

t-Test

Table 4: t-Test: Paired Two Sample for Means

	Unit Price	Total Price
Mean	578.6319	3180.133418
Variance	97515.1	6476916.456
Observations	20000	20000
Pearson Correlation	0.673951	
Hypothesized Mean Difference	0	
Df	19999	
t Stat	-156.831	
P(T<=t) one-tail	0	
t Critical one-tail	1.64493	
P(T<=t) two-tail	0	
t Critical two-tail	1.960083	

This study used the paired t-test to analyze the difference between the means of the unit price and the total price in order to identify the difference. The mean unit price and mean total unit price are 578.63 and 3180.13, respectively with variance 97,515.1 and 6,476,916.46. These major differences may be attributed to the core fact that the price of an individual item substantially differs with the aggregate price with multiple items added together. The number of observations is 20,000 which makes the sample size is large; therefore, they are confident of the results so obtained. When focusing on the dependency of the costs the Pearson correlation is 0.674 which means there is a moderately strong positive relationship with the unit price. This goes a long way to imply that any increase in the unit price is likely to be accompanied by an equal corresponding increase in the total price, consequently underlining how strong a role unit price plays in any firm's revenues. With reference to the hypothesis testing, the hypothesized mean difference is assumed to be equal to zero to stage the null hypothesis whereby, there is no insignificant difference between the mean of unit price and total price. The t-statistic value = -156.831, which is below the t-critical pointer of 1.64493 (1-tailed) and 1.960083 (2-tailed) t-tests at 95% confidence level. P values for one sample t test for one tailed as well as the two tailed are 0 showing the results are statistically significant. The results presented herein provide a basis for rejecting the null hypothesis, which implies that there is a highly significant difference between the mean unit price and the mean total price. This result invokes expectations because total price in a way reflects the cumulative influence of multiple 'unit prices and quantities' in the data. The difference in the means is apparent suggesting the need to appreciate how unit price moderate's quantity and other factors to derive total price. This information can be of benefit to businesses to look into different aspects such as pricing in an aim at boosting their sources of income. For example, concentrating on products with greater unit value or encouraging purchases by quantity could going have a significant impact on raising total revenues. On the one hand, this strong positive coefficient also implies that the ways in which unit prices are adjusted has to be very sensitive to prevent a possible decline in demand. From this analysis, one realizes that pricing is a major determinant of the total sales volumes. Where unit price is concerned, the data reveals a strong positive association with total price when it comes to revenues, using the same data, mean comparison indicates that total prices are significantly different from one another due to compounded calculations. These results call for future studies on the pricing strategies and customers' consumption patterns to improve the revenue generation approach.

Anova: Single Factor
Table,5: ANOVA:

SUMMARY

Groups	Count	Sum	Average	Variance		
Total Price	20000	63602668	3180.133	6476916		
Unit Price	20000	11572637	578.6319	97515.1		
Quantity	20000	109711	5.48555	8.241803		
Add-on Total	20000	1244897	62.24485	3370.781		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1.36E+11	3	4.53E+10	27532.77	0	2.60502
Within Groups	1.32E+11	79996	1644453			
Total	2.67E+11	79999				

The ANOVA results provide insights into the variance between and within groups for the selected variables: Total Price, the price per unit, Total Quantity as well as the amount that the customer has to pay over and above the total price. This means that the frequency distributions of all four groups are identical at 20,000 each, which make it easier to compare observations within the data set. The average values of each of the groups look very dissimilar, the mean value of Total Price is 3180.13, whereas the Quantity is just 5.49, which means the nature of the data is totally different; Total Price is the monetary value of prices, quantity is the count of items bought. The variances within each group also vary: Total Price has the largest variance of 6,476,916, which means that the range of transaction value is broader in this case than in other cases. The statistics in the ANOVA table includes the partitioning of the total variance in to Between Groups and Within Groups. The Between Groups SS is 1.36×10^{11} showing that substantial variability in the data is explained by differences among the variables. On the other hand, the ''Within Groups'' sum of squares, 1.32×10^{11}, explains variability within each variable. The mean square ms for between groups is much higher at 4.53 x 10^10 while the ms for within groups is 1644453 hence the level of variance caused by group differences is much more than that of random variation. The obtained F-statistic of 27,532.77 is very high compared to the F crit of 2.605 sig 0.05 and the obtained P-value of nearly zero. This is evidence that mean differences across Total Price, Unit Price, Quantity and Add-on Total archives a statistical significance. Consequently, there is understanding that these variables represent different phenomena of customer behavior and the dynamics of transactions, which should be addressed methodologically differently. Such findings give a good starting point in future research on these variables as well as their impact on total end sale and buying behaviours of customers.

Regression Analysis

Table 6: Summary Output

Regression Statistics	
Multiple R	0.935887
R Square	0.875885
Adjusted R Square	0.875866
Standard Error	896.6622
Observations	20000

ANOVA

	df	SS	MS	F	Significance F
Regression	3	1.13E+11	3.78E+10	47037.54	0
Residual	19996	1.61E+10	804003.2		
Total	19999	1.3E+11			

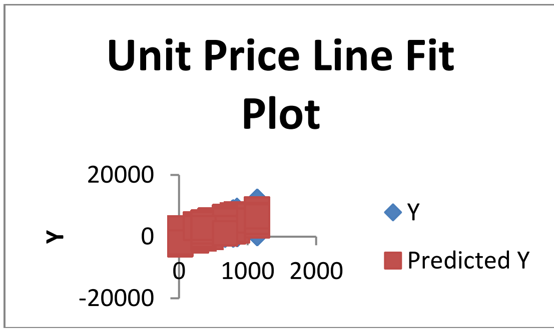
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-3130.74	18.75119	-166.962	0	-3167.494774	-3093.99	-3167.49	-3093.99
Unit Price	5.459248	0.020466	266.7504	0	5.419132991	5.499362	5.419133	5.499362
Quantity	575.669	2.20864	260.6442	0	571.3399328	579.9982	571.3399	579.9982
Add-on Total	-0.09449	0.110076	-0.85843	0.390665	-0.310249518	0.121265	-0.31025	0.121265

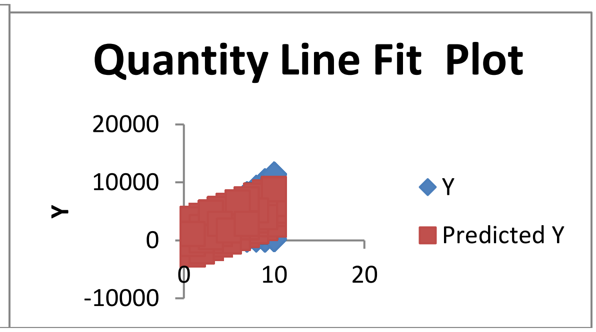
The findings shown by the regression analysis signify a high degree of influence of the independent variables on the measure of the dependent variable, Total Price. The Multiple R value of 0.9359 is the testimony of the fact that there exists high level of correlation existing between the observed values and predicted values in the given model. The R Square of 0.8759 implies that at least 87.59% of the fluctuations in Total Price can be made understandable through the inherent variables of the model. The Adjusted R Square is equally high at 0.8759, this give more credibility to the model given the number of predictors and observations used in developing it. The Standard Error of 896.66 means here average deviation of the observed values from the regression line Interestingly it is not very high way beyond the scale of Total Price. Also, the ANOVA table added strength to the model's significance. It is quite large; thus, meaning the overall regression model is statistically significant overall at the .05 alpha level. The Significance F value is close to zero Another extreme and precious information is the ''Significance F''. This further assures the fact that at least one of the predictors has a significant relationship with the Total Price. The global summary of squares (SS) is 1.3×10^{11}, of which the regression model captured 1.13×10^{11}, while the residuals account for 1.61×10^{10}. We get a large value of the MS for regression compared to the residual MS, emphasizing the ability of the model to explain the variation.

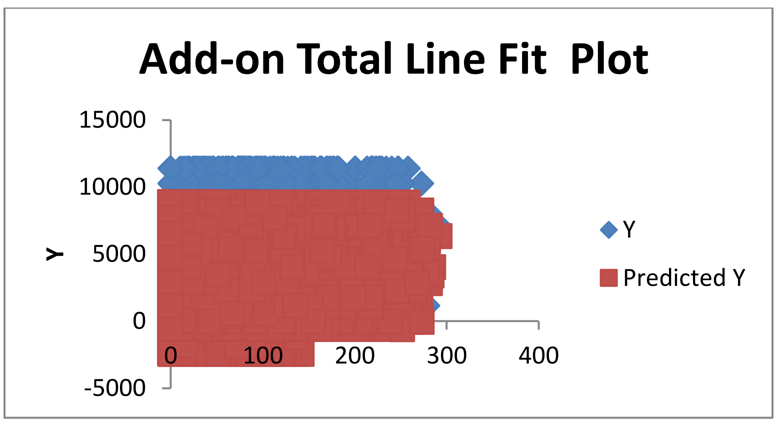
The regression coefficients tell about the contribution made by each predictor. The Intercept equals to -3130.74 which is again meaningful as this is the Total Price when all predictors are nil which can't be the case in this data set. The Unit Price coefficient of 5.459 implies that for an increase of one on the Unit price the total price is likely to increase by 5.459 all else equal. Again, this coefficient is significant at 1% based on both the t-statistic of 266.75 and the P-value that is equal to zero. Likewise, the Quantity coefficient of 575.67 means, each additional unit bought raises the Total Price by 575.67, very close to perfect at incredibly significant level (t-statistic = 260.64; P-value = 0). On the other hand, the Add-on

Total coefficient of -0.0945 is not statistically significance since a significance level of 0.3907 is higher the cut of 0.05. This was an indication that fluctuations in Add-on Total have little effect on the Total Price given the presence of the other variables. Returning to the full model results, confidence interval of significant predictors like the unit price and quantity are relatively tight, indicating precise estimates while the confidence interval of add-on total is wider as for its non-significance. In general, this has further reinforced the fact that Unit Price and Quantity are the most influential predictors of Total Price as opposed to Add-on Total which seems to have very little effect on this model.

Graphical Representation







The three-line fit plots for Unit Price, Quantity, and Add-on Total show actual observed value and predicted value based on the regression model. In the plot: Unit Price Line Fit, the positioning of the predicted values marked by red bars and the actual values marked by blue diamonds indicates a fairly good prediction of the model for the unit price. The distribution of points is increasing as the unit price increases, and there are scatter points in the higher price range; the limitations of the prediction precision are more significant at lower unit price values. For the Quantity Line Fit Plot, the red as well as the blue markers are inclined upwards and this indicates that the quantity and the dependent variable moves in the right direction. This consistency assures the reliability of estimate based on quantity key area hence leading to an accurate estimation of sales as depicted below. The link is somewhat narrow in this case, which implies a good level of the predictive accuracy of quantity. The Add-on Total Line Fit Plot also shows a larger scatter between the predicted and actual values especially with relatively high values of add-on totals. There is some level of corresponding in values for lower add-on totals; however, the fluctuation escalates as values go up, meaning that the model is slightly off with the estimations of sales when the add-on totals are higher. This also implies that add-on total holds lesser of predictive ability as compared to the unit price and quantity. Both these plots show the general performance of the regression model in relation to the total variance with the variation precision across the predictors.

Conclusion and Discussion

In summary, the paper's e-commerce sales data insights have given an understanding of the potential factors that motivate or lead to customer purchase decisions. Similar to the previous analysis, the regression conducted for unit price and quantity showed both are good predictors of total sales having a high t-stat and being statistically significant at 1%. A measure consistent with this was the coefficient of determination of the total sales that revealed unit price as the variable with the highest predictive power and this was positively signed to suggest that generally, products that are sold at higher unit prices were likely to make higher contributions to total sales. As with the number of units bought, there was a very strong and significant positive correlation between the quantity bought and total sales, again making a point that a larger quantity purchased significantly contributes to total sales. However, concerning the add-on total, there was shown a low and unstable relation with the sales figures, which means additional purchases or accessory items have a low effect on the gross figure. This suggests an opportunity for further enhancement in the area of marketing or packaging strategies for additional or complimentary products. The descriptive statistics also emphasize the dispersion of customer spending behavior. The variability in total price and especially in the unit price are very large, reflected by a large standard deviation, which means that the customers and their buying capacities are rather diverse. The results revealed that both sales and unit prices have positive skewness and excess kurtosis, which indicate that this variable is not normally distributed and is leaned towards more on the right side, which shows that most customer spends less than some customers spend more. Also, the ANOVA analysis establishes the significance of the difference between variable stores with a high F-value that substantiates the predictiveness of the predictors. Hence, these findings support calls for the application of a marketing mix to target different customers and provide the right price and available quantity to offer for specific products to ensure high returns. However, a useful written conversation can be performed while appreciating some limitations of the analysis carried out in the study. First, the results of the regression model show residual by means of the R-squared value, which equals 0.876, meaning that 12.4% of variance in a total amount of sales is unaccounted for. They think that there are other factors that might affect sales, including customer characteristics or seasonal characteristics that were not considered in this study. Moreover, while using the line fit plots, especially for add-on total, some deviations between actual and forecasted quantities have been revealed, which point to the potential enhancement of the predictive capacity of the model. The poor predictiveness of add-on totals means that we have to consider that research may be needed with additional information or by application of different algorithms. A limitation that is obviously derived from the method used in this study arises from the static nature of the data collected during the analysis. At the same time, the dataset is specific by time, which means that customer behavior can change over time thanks to economic, cultural, or market factors. Thus, the findings may not extend to other temporal or spatial settings. Furthermore, the evaluation of the model is carried out based on numerical indicators, not qualitative aspects like customer satisfaction or loyalty or any kind of product reviews that may explain purchasing behavior in detail. It means that using such dimensions in future research may provide additional knowledge of the matter. Finally, the breakdown of results concerning unit price and quantity as the leading factors contributing to e-commerce sales suggests that there is much to be learned about optimizing add-on totals. The research proposes recommendations for managers in e-commerce companies with a special focus on the strategic level of pricing and bulk purchasing. However, it also emerges that there exists a need for more elaborate models incorporating other factors and trends over time. Other research ideas that may be of interest in the future include the use of customer-oriented measures or the effects of variations across the seasons of the year on purchasing activity or the utilization of more sophisticated machine learning approaches in order to increase the accuracy of the various predictions and better account for the multifaceted nature of customer behavior. If these limitations are countered, it will be easy for the business to factor in the new environments of e-commerce and come up with new strategies that will correspond to the new market needs and desires.

References

Biswas, A., Chattopadhyay, S., Barai, P. & Samanta, P. 2024. Understanding the relationship between in-store and online shopping channel choice behaviour of customers: a systematic literature review. Cogent Business & Management, 11**,** 2402510.

BLACK, J. J. 2024. Predictors of Online Purchase Conversions Using Clickstream Data and Sentiment Analysis. University of South Alabama.

Chen, Y., Li, M., Song, J., Ma, X., Jiang, Y., Wu, S., & Chen, G. L. (2022). A study of cross-border E-commerce research trends: Based on knowledge mapping and literature analysis. Frontiers in Psychology, 13, 1009216.

Du, W., &Dzamankulov, B. (2023). The value, role, and analysis of cross-border e-commerce development in china's import and export economy. International Journal of eBusiness and eGovernment Studies, 15(1), 366-393.

FAIZA, F. & TAHER, K. A. Consumer Insights in E-commerce: Analyzing Sales Data Using Clustering Algorithm. 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), 2024. IEEE, 670-674.

Hashemian, F., Maleki, N. & Zeinali, Y. 2024. From User Behavior to Subscription Sales: An Insight Into E-Book Platform Leveraging Customer Segmentation and A/B Testing. Services Marketing Quarterly, 45**,** 153-181.

Jin, L., & Chen, L. (2024). Exploring the Impact of Computer Applications on Cross-Border E-commerce Performance. IEEE Access.

Khuan, H., Rahmiyati, N., Mendrofa, K. J., Diwyarthi, N. D. M. S. & Wiartha, N. G. M. 2024. Analysis of Consumer Perceptions of Product Quality, Sales Promotion, and Ease of Purchase in Increasing Purchasing Decisions at Start-up Companies in Indonesia. West Science Interdisciplinary Studies, 2**,** 264-271.

Lu, H., & Liang, J. (2024, August). Consumer Behavior Analysis Model of Cross-border E-commerce Based on Big Data. In 2024 International Conference on Power, Electrical Engineering, Electronics and Control (PEEEC) (pp. 722-726). IEEE.

Mou, J., Ren, G., Qin, C., & Kurcz, K. (2019). Understanding the topics of export cross-border e-commerce consumers feedback: an LDA approach. Electronic Commerce Research, 19, 749-777.

MADYATMADJA, E. D., AMENDI, R., GIGGS, D. R. & TANADY, F. Exploring the Impact of Social Media Shopping Features on Student Buying Behavior: A Study in Indonesia. 2024 10th International Conference on Smart Computing and Communication (ICSCC), 2024. IEEE, 274-280.

Raji, M. A., Olodo, H. B., Oke, T. T., Addy, W. A., Ofodile, O. C. & Oyewole, A. T. 2024. E-commerce and consumer behavior: A review of AI-powered personalization and market trends. GSC Advanced Research and Reviews, 18**,** 066-077.

Sunarya, P. A., Rahardja, U., Chen, S. C., Lic, Y.-M. & Hardini, M. 2024. Deciphering digital social dynamics: A comparative study of logistic regression and random forest in predicting e-commerce customer behavior. Journal of Applied Data Sciences, 5**,** 100-113.

Tarnanidis, T. 2024. Exploring the Impact of Mobile Marketing Strategies on Consumer Behavior: A Comprehensive Analysis. International Journal of Information, Business and Management, 16**,** 1-17.

Tripathi, V. V. R., Srivastava, M. K., Jaiswal, R., Singh, T. D. & Khaled, A. S. 2024. Marketing logistics and consumer behaviour: an empirical study on Indian e-shoppers. Cogent Business & Management, 11**,** 2397559.

Tang, W., & Li, G. (2023). Enhancing competitiveness in cross-border e-commerce through knowledge-based consumer perception theory: an exploration of translation ability. Journal of the Knowledge Economy, 1-34.

Wang, C. (2021). Analyzing the Effects of Cross-Border E-Commerce Industry Transfer Using Big Data. Mobile Information Systems, 2021(1), 9916304.

Yan, L. (2022). [Retracted] Predictive Analysis of User Behavior Processes in Cross-Border E-Commerce Enterprises Based on Deep Learning Models. Security and Communication Networks, 2022(1), 1560017.

Yang, H., Luo, Y., Qiu, Y., Zou, J., Masukujjaman, M. & Ibrahim, A. M. 2023. Modeling the enablers of consumers' E-shopping behavior: A multi-analytic approach. Sustainability, 15**,** 6564.

ZHANG, S. 2024. The role of Artificial Intelligence in enhancing online sales and the customer experience.

Zhang, X., Xu, D., & Xiao, L. (2021). Intelligent Perception System of Big Data Decision in Cross-Border e-Commerce Based on Data Fusion. Journal of Sensors, 2021(1), 7021151.