E-Commerce Sales Data Analysis Report

By

Odunola Akimu

Data Science Intern

@ ArchSaint Nexus Camp

03/02/2026

1. Introduction

This report presents a comprehensive exploratory and revenue-focused analysis of an e-commerce sales dataset. The primary objective is to understand data structure, quality, and key business patterns related to customer purchasing behavior, product performance, revenue generation, refunds, regional trends, payment methods, and the impact of discounts. Insights from this analysis provide a solid foundation for strategic business and pricing decisions.

2. Dataset Description

The dataset contains 3,000 customer order records with 9 key features covering **order dates, product categories, order status, payment methods, regional information, discounts, and revenue**. Initial data inspection confirmed the absence of duplicate records. The **order_date** variable was originally stored in an incorrect format and was converted into a proper date format to enable temporal and seasonal analysis.

3. Data Cleaning and Preprocessing

Missing values were identified in the **discount_applied** column, accounting for 610 observations. To maintain realistic purchasing behavior and avoid regional bias, missing values were imputed using the median discount value within each region. This approach preserved local pricing patterns and ensured consistency across regions.

4. Feature Engineering

To support temporal and seasonality analysis, additional features were extracted from the order date, including weekday, month name, month number, day, week number, and a weekend indicator. The year feature was removed since all records belonged to a single year (2024), offering no additional analytical value.

5. Categorical Feature Analysis

5.1 Product Categories

The dataset includes five product categories: Groceries, Fashion, Electronics, Home, and Beauty. Fashion recorded the highest order volume with 638 transactions, while Groceries recorded the lowest with 554 transactions.

5.2 Payment Methods

Four payment methods were observed: Bank Transfer, Wallet, Card, and USSD. Bank Transfer was the most frequently used method with 763 transactions, while USSD recorded the lowest usage with 722 transactions.

5.3 Order Status

Order outcomes were classified as completed, cancelled, or refunded. Completed orders were the most common with 1,025 transactions, followed by cancelled orders (995) and refunded orders (980).

5.4 Regional Distribution

The dataset spans 50 regions. Connecticut recorded the highest order volume with 77 transactions, while regions such as Kansas and North Dakota recorded the lowest order counts.

6. Temporal and Seasonality Analysis

Customer activity was higher on weekdays, accounting for 2,133 orders, compared to 867 orders on weekends. Monthly analysis showed peak activity in April (272 orders) and the lowest activity in September (236 orders), indicating seasonal variations in purchasing behavior.

7. Numerical Feature Assessment

Exploratory analysis using boxplots revealed no significant outliers in numerical features, indicating stable distributions suitable for further statistical analysis and modeling.

8. Revenue Analysis for Completed Orders

Home products generated the highest total revenue from completed orders at $56,116.08, making them the top-performing category in terms of successful sales. Groceries recorded the lowest completed-order revenue at $50,079.83. Beauty products achieved the highest average revenue per completed order ($272.62), while Fashion recorded the lowest ($243.07).

9. Refund and Revenue Loss Analysis

Refund analysis identified Fashion as the largest contributor to refunded revenue, resulting in losses of $57,563.58. Groceries followed with $41,591.24 in refunded revenue. Electronics recorded the lowest average refunded revenue per order ($266.29), indicating relatively lower financial risk per refund.

10. Regional Revenue Performance

Wyoming generated the highest total revenue at $8,095.00, while Indiana recorded the lowest at $2,684.51. In terms of refunded revenue, Washington recorded the highest loss ($8,373.46), while Iowa recorded the lowest ($2,804.05).

11. Payment Method Performance

Bank Transfer emerged as the most effective payment method for completed orders, generating $74,047.54 from 275 successful transactions. USSD recorded the lowest completed-order revenue ($55,674.57). Wallet payments accounted for the highest number of refunded transactions (258), resulting in a revenue loss of $67,857.08.

12. Discount Analysis

12.1 Completed Transactions Without Discounts

Groceries recorded the highest number of successful zero-discount transactions, while Home recorded the lowest. Card payments dominated zero-discount completed transactions, generating the highest revenue. Regionally, Colorado generated the highest zero-discount revenue, while New Mexico recorded the lowest.

12.2 Completed Transactions With Discounts

Discounted transactions significantly increased sales volume and revenue, particularly for Home products, which generated $49,868.83. Bank Transfer was the most used payment method under discounted completed transactions. Discounted sales were strongest during weekdays and peaked in January.

12.3 Refunded Transactions Without Discounts

Home products recorded the highest revenue loss among non-discount refunds. Bank Transfer resulted in the highest refunded revenue despite fewer transactions. January recorded the highest refund losses, while August recorded the lowest.

12.4 Refunded Transactions With Discounts

Fashion products recorded the highest refunded revenue under discounted conditions ($48,399.44). Wallet and Card payments contributed significantly to refunded revenue. Regionally, Washington recorded the highest discounted refund losses.

13. Summary of Findings

The dataset is clean, well-structured, and suitable for advanced analysis. Home and Beauty products perform strongly in completed sales, while Fashion and Groceries contribute disproportionately to refund-related losses. Discounts increase transaction volume and revenue but also elevate refund risks, particularly for Fashion products and Wallet payments.

14. Business Recommendations

The business should prioritize Home and Beauty products through targeted promotions and inventory optimization. Refund-heavy categories such as Fashion and Groceries require improved quality control, clearer product descriptions, and better logistics. Bank Transfer should be promoted as a preferred payment option due to its strong performance and lower refund rates. Weekday-focused marketing campaigns and region-specific strategies are recommended to maximize revenue and reduce losses.

15. Discount Optimization and Risk-Control Strategy

Discounts should be applied selectively based on product risk and performance. Home products are suitable for moderate discount expansion, while Fashion discounts should be controlled and targeted. Tiered discount structures, payment-method-specific incentives, region-based promotions, and continuous monitoring through dashboards are recommended to balance revenue growth with refund risk.

16. Conclusion

This analysis demonstrates that data-driven discount and pricing strategies can significantly improve e-commerce performance. By aligning discounts with product behavior, payment reliability, regional trends, and timing, the business can increase profitability, reduce refund-related losses, and support sustainable long-term growth.