Amazon Sales — SQL Data Analysis Report

Author: Dawid Jasiński

Technologies: SQL Server, SSMS

Date: 18.02.2026

Executive Summary

The analytical work conducted on the e-commerce sales data covering the years 2022–2023 enabled the team to perform a comprehensive assessment of the revenue structure, the effectiveness of the discount policy, seasonality patterns, and key operational performance indicators. Leveraging SQL Server as the primary analytical environment made it possible to construct a normalized relational model that served as the foundation for all subsequent analyses. The initial CSV dataset was transformed into a three-layer structure consisting of the Orders, Products and Order_Items tables, providing an accurate representation of relationships between orders, products and individual sales items. This normalized architecture eliminated redundancy and ensured alignment with data management best practices.

The analysis of product categories revealed that sales distribution across segments is relatively balanced, indicating a broad and diversified product portfolio. The Beauty category recorded the highest revenue in the analyzed period (5.55M), while Sports generated the lowest result (5.40M). The modest variance between categories suggests that sales performance is not heavily dependent on any single segment, which in the long term reduces the risk of revenue concentration.

A similarly even distribution was observed in the regional analysis. Each of the four markets—Middle East, North America, Asia and Europe—generated revenue exceeding 8M, with the Middle East achieving a slightly stronger outcome (8.30M). Europe, despite being the lowest of the four, remains a region with significant potential. This geographic balance indicates a stable international presence, reducing operational risk through diversification across multiple markets.

From a temporal perspective, sales exhibited a high degree of stability with identifiable seasonal peaks and troughs. The strongest monthly performance was recorded in January 2023 (1.46M), while the weakest occurred in February 2023 (1.24M). These fluctuations may reflect year-end and post-holiday behaviours, as well as the impact of promotional

campaigns typically associated with these periods. Confirming these observations would require additional analysis of marketing activities and promotional calendars.

The assessment of discount effectiveness proved to be one of the most insightful elements of the study. The team observed that the highest overall revenue was generated at discount levels between 0–5%, while increasing discounts up to 30% resulted in a pronounced decline in both total revenue and the average order value. The findings clearly indicate that aggressive discounting does not lead to higher transaction volumes and instead reduces the value of orders—ultimately lowering the profitability of the platform.

Key performance indicators further reinforced the platform's stable operating profile. Total revenue reached 32.87M, with 50,000 processed orders. The average order value amounted to approximately 657 units, while customers typically purchased two products per transaction. These results suggest that the platform focuses on mid-priced products that are purchased in small but consistent quantities.

Collectively, the analysis shows that the platform maintains a stable level of sales with limited volatility across categories and regions. The discount policy requires optimization toward lower discount ranges, which proved to be the most effective in driving revenue. Additionally, recognizing recurring seasonal patterns presents opportunities for more targeted marketing and promotional strategies. The results of the analysis form a solid foundation for further work, including dashboard development and extended analytics supporting decision-making in sales and marketing.

.

Introduction and Project Objectives

As part of the analytical project, the team focused on a comprehensive examination of sales data originating from a major e-commerce platform, covering the period from January 2022 to December 2023. The objective of the analysis was not limited to describing current performance; it was primarily aimed at identifying underlying patterns and relationships that could form the basis for informed decision-making at both operational and strategic levels. The project aligns with practices common in large organizations, where continuous improvements in data management—through integration, cleansing and modeling—serve as the foundation for developing advanced reporting, controlling and predictive analytics capabilities.

The initial dataset, obtained from an external source, had a transactional structure and contained detailed information regarding orders, products, customer ratings, and financial metrics representing the final value of each transaction. Such datasets are typical starting points for sales analysis across many retail and e-commerce organizations, yet their direct use in analytical environments is often challenging due to repetition, lack of hierarchical structure and inconsistent data quality across fields. In line with standard practices applied in professional consulting and enterprise analytics environments, the team began by constructing a coherent, normalized data model that clearly separated order-level information from product-level attributes and transaction line-item details. This separation improved analytical clarity and ensured consistency with corporate-grade relational data modeling principles.

The core purpose of the project was to address key questions related to sales dynamics, revenue structure and customer purchasing behavior. Of particular interest were differences in revenue across product categories, sales performance across geographic regions and seasonal patterns that might reveal periods of increased or decreased customer activity. An equally important aspect involved evaluating the effectiveness of the discount policy, given that discounting is one of the most frequently used levers for influencing sales volume, while at the same time being a major determinant of profitability. The team aimed to determine the extent to which discounts actually translate into revenue growth and whether certain discount levels are disproportionately inefficient.

The project also included an assessment of key performance indicators (KPIs), such as total sales volume, number of orders, average order value and the average number of items purchased per transaction. These metrics are central to evaluating the operational performance of any e-commerce organization, and interpreting them correctly enables a deeper understanding of customer behavior and product portfolio characteristics. Additionally, the project aimed to formulate conclusions and recommendations that could support decision-making across sales, marketing and promotional strategy domains.

Consistent with consulting projects executed for corporate clients, the analysis combined both quantitative and qualitative elements. The team's priority was to ensure that all insights derived directly from the data while presenting them in a manner accessible to business stakeholders—regardless of their technical background. The goal was to create a document that could serve both as a standalone analytical report and as a foundation for subsequent work, including business intelligence dashboards, extended statistical analysis or the

Data Set Description and Analysis Methodology

The dataset used in this analysis was provided in the form of a CSV file containing the full transaction history of an e-commerce platform covering the period from January 2022 to December 2023. The data was transactional in nature and included both order-level information and details related to individual product items, along with attributes such as category, rating, and number of reviews. In its original form, the dataset had a flat structure, meaning that a single table simultaneously contained information about orders, products, and revenue-related metrics derived from them. This format is typical for most CSV-based datasets; however, in the context of professional business analytics and reporting system design, it does not represent an optimal data structure. For this reason, one of the first tasks of the team was to perform data normalization, allowing the separation of information according to natural relationships and reducing redundancy.

During the preprocessing stage, the data was imported into a SQL Server environment and transformed into a three-layer relational model. The central component of the model was the Orders table, which included order identifiers, purchase dates, customer regions, and applied payment methods. The second table, Products, was purposely designed in a simplified form and contained only unique product identifiers. This design decision stemmed from the observation that certain attributes—such as product rating or category—were not stable across records and depended more on the transactional context than the product itself. Consequently, these attributes were retained in the third table, Order_Items, which served as the main fact table containing purchased items, prices, discounts, quantities, categories, and revenue values associated with each transaction.

This structure enabled conducting analyses in a manner consistent with standards used in data-warehouse design and management reporting. The relational model improved clarity, ensured consistency of relationships within the dataset, and increased the efficiency of analytical queries. Additionally, normalization made it possible to clearly separate elements that belong to the order level from those associated with individual line items, thus allowing analyses that capture both product-level trends and customer behavior while avoiding information duplication that could distort interpretation.

After completing the data-modeling stage, the team proceeded to validation and sanity-checking. This step involved verifying numerical values, identifying potential duplicates, and assessing the consistency of order dates. The dataset size was also checked, confirming the presence of 50,000 unique orders and more than 32 million revenue units. These activities reflect the standard data-quality procedures used in organizations, where all analytical data must be audited prior to interpretation.

Once validation was complete, the team developed a set of SQL queries designed to address the research questions defined at the start of the project. These queries ranged

from simple aggregations to more advanced analyses involving table joins, window functions, and grouping operations. Their purpose was to examine revenue distribution across categories, regions, and time periods, as well as to evaluate the impact of discounts on customer purchasing behavior. This approach reflects analytical practices typically used in corporate environments, where the logical correctness, clarity, and reusability of queries—for example in reporting or dashboard development—are of key importance.

A crucial methodological element was the development of a set of KPI indicators. These metrics—including average order value, total revenue, number of orders, and average number of units purchased per order—were computed directly from the normalized dataset. This ensured more reliable results and eliminated the risk of double-counting or distorted outputs caused by duplicated information in the raw data. Finally, all analytical steps were integrated into a comprehensive business narrative, explaining the results, their interpretation, and the potential implications for strategic decision-making in the areas of sales and marketing.

Characteristics of the Normalized Data Model

A key stage of the project involved building a normalized data model that could serve as a foundation for reliable and repeatable sales analysis. This approach aligns with practices commonly used in large organizations, where the priority is to ensure maximum data transparency, eliminate redundancy, and guarantee unambiguous relationships between individual dataset elements. In the case of the analyzed dataset, the main challenge was its initial structure, which combined order-level information, product details, and product attributes within a single flat table. Although such structures are common in file-based environments and exploratory projects, they are insufficient—and sometimes problematic—for more advanced business analytics, as they hinder cross-sectional analysis and the creation of operational reporting later on.

The analytical team decided to transform the original dataset into a model composed of three tables: Orders, Products, and Order_Items. This decision was driven by an examination of the dataset's structure, which revealed numerous repeated values at the level of individual transactions. For example, the order identifier reappeared with every product row, and product identifiers were repeated thousands of times depending on how many orders included a given item. In corporate data environments, this leads to unnecessary system load as well as increased risk of analytical errors. A normalized design eliminates these issues.

The Orders table was designed as a collection of attributes describing each order at the general level. It contains only those data points that remain constant within a transaction: the order ID, order date, customer region, and payment method. Separating these attributes from other information was necessary to create a clear and logically consistent model in which each order appears exactly once, regardless of how many product items it contains. This design follows relational-database principles consistent with Third Normal Form.

The second component of the model, the Products table, was intentionally simplified. In the analyzed dataset, details such as ratings or product categories were not stable product-level characteristics but varied at the level of individual transactions. This suggests that these attributes were not inherent properties of the product itself but were tied to the sales context—likely reflecting different points in time, data sources, or product catalog variations. Therefore, the team intentionally limited the Products table to product identifiers only, removing inconsistent attributes that could otherwise lead to misinterpretation. Such a design is often used when data originates from multiple transactional systems or external sources that do not guarantee attribute consistency.

The third component of the model, the Order_Items table, functions as the fact table, containing details for every item purchased within each order. This is where information such as product ID, unit price, discount level, number of units, item category, rating, and

the final revenue generated by each line item is stored. In practice, this table forms the basis of most sales analyses, as it captures both quantitative and value-based aspects of transactions. Its relationship with the Orders table enables analyses by region or purchase date, while its relationship with the Products table supports product-level and SKU-level examination.

This data-modeling approach supports analytics aligned with best practices in Business Intelligence and Data Analytics. It prevents data redundancy that could distort aggregated results or cause double-counting of revenue. Furthermore, the three-table structure is fully compatible with data-warehouse design techniques, including star and snowflake schema modeling. This means the model can be expanded in the future with additional dimensions—such as customer segment, marketing campaign, or sales channel—without requiring architectural redesign.

Finally, it is worth emphasizing that the applied normalization plays a critical role in ensuring the quality of the analysis. Separating order-level data from line-item and product-related details significantly reduces the risk of errors stemming from inconsistencies in the dataset, while also enabling more efficient creation of analytical views, stored procedures, and queries used in later project stages. In practice, this means the model not only supports current analytical work but also creates a solid foundation for future reporting tools, including Power BI dashboards and predictive-analytics modules.

Sales Analysis by Product Category

As part of the sales analysis, one of the key focus areas was evaluating the structure of revenues generated by individual product categories. In the context of e-commerce operations, this type of analysis is fundamental both for monitoring the performance of different segments of the product offering and for identifying categories with the highest growth potential or those requiring further optimization. The conducted analysis revealed that the sales structure during the examined period was characterized by a high degree of balance across categories, indicating a broad and diversified product portfolio. From a business perspective, such distribution can be advantageous, as it reduces the risk of financial dependence on any single dominant segment.

During the analyzed period, the highest revenue level was generated by the Beauty category, which reached a total of 5.55 million in sales value. Although this result placed Beauty as the top-performing segment, it did not differ significantly from the values observed in other categories. The Books and Fashion categories both achieved 5.48 million, while Home & Kitchen and Electronics each generated 5.47 million. The lowest revenue level was observed in the Sports category, with 5.40 million. The difference between the strongest and weakest category amounted to only about 150,000—a minor gap considering the overall revenue scale of the platform.

Such a sales distribution suggests that platform users exhibit a relatively wide range of purchasing interests, and the product offering is designed to be evenly diversified rather than concentrated on a single dominant product type. This stands in contrast to many other e-commerce platforms, where one or two categories—such as electronics or home products—typically account for a disproportionate share of total revenue. In this case, the observed balance may result both from the nature of the dataset and from the design of the product offering, which is likely positioned to appeal to a broad and universal audience.

From a business standpoint, an evenly distributed revenue mix across categories can be interpreted in two ways. On one hand, it is a positive indicator of stability, suggesting that even if demand declines for a specific product type, the platform is unlikely to experience significant revenue fluctuations. On the other hand, the absence of clear category leaders may indicate a relatively low degree of specialization or a lack of segments with distinctly higher profitability. In practice, this suggests that the platform's strategy likely focuses on broad product availability rather than dominance within a particular niche.

It is also noteworthy that five out of the six categories fall within a very narrow revenue range. This may indicate that marketing activities, promotional campaigns, and customer-targeted initiatives are similarly effective regardless of product category. From a business perspective, this could reflect a balanced marketing budget or an absence of

strong prioritization of specific categories. Nevertheless, the slight advantages observed for the Beauty and Books categories may justify deeper analysis that considers margins and each category's contribution to customer conversion. Revenue alone does not provide a complete picture of financial performance—equally important are acquisition costs, margins, and levels of returns and complaints.

In the context of offer management, the results suggest that the platform should consider conducting additional evaluations not only in terms of revenue but also profitability and growth dynamics. For instance, if the Beauty category—despite its modest revenue lead—also achieves higher margins or shows lower sensitivity to seasonality, it could represent a more strategically attractive investment area. Conversely, the Sports category, despite generating the lowest revenue, may still be valuable if it produces strong margins or appeals to a highly loyal customer segment.

The analytical team also noted that the balanced category distribution contributes to long-term revenue stability. However, to fully leverage the potential of each category, it is recommended to perform complementary analyses, such as examining user behavior, shopping basket composition, and purchase frequency by category. In practice, modern e-commerce platforms increasingly rely on such insights to support customer segmentation and product personalization, which can enhance conversion rates and increase revenue per user.

In summary, the analysis of sales by product category indicates a highly balanced distribution of revenues across the product offering. This provides a solid foundation for further scaling of operations. At the same time, the results serve as a starting point for deeper profitability assessments and cost-efficiency analyses for each category, supporting strategic decision-making in product management, marketing budget allocation, and long-term planning.

Sales Analysis by Region

The evaluation of sales performance across geographic regions is a key component of business analysis, as it allows organizations to determine which markets generate the highest revenue volumes and how consumer behavior varies depending on location. From the perspective of sales management and marketing planning, such information plays a crucial role, enabling the identification of strategic areas that require increased attention as well as stable regions that support the platform's long-term and predictable performance. In the analyzed dataset, sales were divided into four primary markets: Middle East, North America, Asia, and Europe, representing a typical segmentation structure used in reporting for global e-commerce platforms.

The results of the analysis indicate that each of the four regions plays a significant role in generating revenue, with sales values falling within a relatively narrow range. The highest revenue was recorded in the Middle East, which generated 8.30 million units of sales value. Only slightly lower results were achieved in North America at 8.28 million. The next positions were Asia with 8.17 million and Europe, which generated 8.11 million. The difference between the strongest and weakest market amounted to roughly 190,000—a small value relative to the platform's total revenue scale and evidence of an exceptionally balanced distribution of sales across regions.

Such a distribution is not typical in the e-commerce industry, where one dominant market often drives the majority of revenues. For many global platforms, the North American or European market accounts for the largest share, while eastern markets tend to represent smaller supplementary segments. The dataset examined here presents a different situation: each region serves as an equally important pillar of overall sales. This suggests that the platform likely operates in a multi-regional model, with its offering, logistics, and marketing activities optimized to deliver consistent performance across all analyzed markets.

This revenue balance may constitute a significant competitive advantage. Platforms dependent on a single market are more vulnerable to local economic fluctuations, regulatory changes, or seasonal declines in demand. In this case, such risks are minimized, as no single region determines the overall condition of the platform's sales. Therefore, it can be concluded that the platform exhibits high resilience to geographic risk. Regardless of changes occurring in one area, the remaining regions maintain stable performance, which supports long-term predictability of results.

However, the regional analysis should not be limited solely to revenue figures. Although revenues remain evenly distributed, important differences may exist in product mix, seasonality patterns, or discount levels applied in each region. While the current analysis focuses primarily on revenue data, potential avenues for deeper exploration are already visible. For example, the Middle East region—currently the top revenue

generator—may exhibit a different basket structure or respond differently to promotional activities than other markets. Similarly, Europe, despite ranking lowest in total sales, may demonstrate stronger customer loyalty or lower sensitivity to economic fluctuations.

The analytical team also noted that the stability observed across regions may suggest a similar level of market penetration and a unified customer-acquisition strategy. If the platform applies comparable marketing budgets and mechanisms across different geographic areas, then the lack of significant revenue discrepancies is a natural consequence. This could imply that advertising efforts and promotional activities have been optimized to achieve consistently high effectiveness, regardless of region. In practice, this may indicate that the platform operates under a universal sales model that performs well across diverse cultural and economic environments.

Interpreting these results should also involve external market context, including the specificity of each region. Purchasing preferences, disposable income levels, and competitive saturation differ across markets. Therefore, the observed revenue balance may stem from both strategic positioning of the offering and customer adaptability. A platform capable of tailoring its assortment to diverse geographic groups typically gains an advantage in scaling operations. In the case of the analyzed dataset, the platform appears to generate stable and comparable results despite these differences.

Overall, the findings highlight the importance of a multi-regional approach to e-commerce sales. Achieving revenue balance across regions is challenging and requires refined logistics as well as a consistent customer-service standard. The results suggest that the platform is well prepared to operate under varied market conditions and can adapt its offering in a way that supports stable revenue generation in each analyzed region. From a business perspective, this provides a solid foundation for further growth, including expansion into additional markets or strengthening competitive advantages within current regions.

Seasonality Analysis

Seasonality analysis is one of the key elements of sales data evaluation, as it allows identification of periods during which demand for offered products increases or declines. For e-commerce platforms, seasonality plays a particularly important role, because cyclical changes in consumer behavior directly affect revenue, inventory requirements, logistics infrastructure load, and promotional strategy. From a business perspective, correctly identifying seasonal patterns enables effective planning of marketing campaigns, inventory management, and decisions regarding the allocation of advertising budgets. In this project, seasonality was assessed by analyzing monthly revenue values for the years 2022–2023.

Monthly data shows that sales remained stable throughout the examined period; however, noticeable differences appeared between individual months. The highest revenue was recorded in January 2023, where sales reached 1.464 million units. This value significantly exceeds those of other months, suggesting that the period immediately following the New Year stimulated increased purchasing activity. This pattern is typical for many online retail platforms, where sales rise due to post-holiday promotions, the use of gift cards received during the Christmas season, and clearance sales of seasonal products.

Conversely, the lowest sales performance occurred in February 2023, when monthly revenue fell to 1.238 million. Such a marked drop following a strong January is a recurring phenomenon in e-commerce. February is a shorter month and lacks major commercial holidays or sales events, resulting in a natural lull in customer activity. Therefore, the February decline can be interpreted as a standard seasonal deviation rather than a signal of deteriorating performance.

A broader two-year perspective reveals additional trends. August 2022 and July–August 2023 stand out with higher sales levels, reaching approximately 1.449 million and 1.442 million respectively. This pattern may be influenced by the summer period, when consumers make purchases related to travel, replace household items after intensive use, or prepare for the back-to-school season. Summer months also typically drive increased activity in categories such as apparel, sports, and home goods, which can significantly impact aggregated platform results.

Reviewing the monthly values also indicates that despite noticeable peaks and declines, overall sales do not exhibit extreme volatility. Most months fall within the range of 1.30–1.44 million, demonstrating strong revenue stability—typical for platforms with broad product portfolios and well-developed customer bases. In practice, this means the platform is not excessively dependent on a single seasonal period but generates consistent revenue throughout the year.

From an analytical perspective, it is also relevant to examine quarterly dynamics. In this dataset, no significant deviations between quarters were observed, reinforcing the platform's stability. The absence of sharp increases in Q4—normally associated with the holiday season—may reflect the characteristics of the dataset, which likely does not include major sales events such as Black Friday. It may also indicate that the platform does not implement large-scale discount campaigns typical for the retail industry, or that such promotions do not substantially impact monthly revenue. Regardless of the underlying cause, the observed data suggests that results are driven more by stable demand than by intense seasonal campaigns.

In terms of business actions, seasonality analysis provides valuable insights for planning pricing strategy and promotional activities. If the platform consistently sees stronger performance in January or August, it may consider increasing marketing investments during periods of naturally higher conversion. Conversely, weaker months such as February represent an opportunity to introduce sales-stimulating initiatives such as educational campaigns, enhanced visibility of selected categories, or moderate discounts—which, according to the discount analysis, are most effective in the 0–5% range. Such measures could help mitigate natural seasonal declines and improve monthly outcomes.

Finally, the seasonality patterns observed in the data form a strong basis for further predictive analysis. Forecasting models built on historical data could be used to project monthly sales levels, facilitating more efficient inventory planning and logistics management. In such models, seasonality plays a pivotal role, and the stability observed in this dataset makes it well-suited for methods such as Holt-Winters, ARIMA, or more advanced machine-learning approaches.

Overall, the seasonality analysis indicates that the platform operates under conditions of moderate and predictable demand fluctuations that do not pose a risk to revenue stability. This is a favorable signal from an operational-management standpoint and highlights potential areas for further optimization in marketing strategy and operational planning.

Discount Policy Effectiveness Analysis

The discount policy is one of the most important tools used by commercial organizations to stimulate demand, increase sales volume, and support inventory rotation. At the same time, it is an area that directly affects profit margins and overall financial performance, since every granted discount reduces unit revenue. For this reason, evaluating the effectiveness of discounts should be conducted systematically and based on data rather than intuition or habitual practices. In this project, the team focused on examining the relationship between discount levels and generated revenue, as well as the average order value. The aim was to determine which discount levels are truly beneficial from a business perspective and which may reduce profitability without producing proportional increases in sales.

The analysis was performed using line-item–level data, allowing precise linkage of each transaction to a specific discount rate. The team grouped the data by percentage discount levels and calculated the number of sold items, total revenue, and average revenue per item for each group. The dataset exhibited six main discount tiers: 0%, 5%, 10%, 15%, 20%, and 30%. This approach enabled an assessment of both sales volume and sales quality in terms of revenue generation.

The results clearly show that the highest total revenue is generated within the low-discount segment. Sales without any discount produced 6.15 million in revenue, with an average item value of approximately 749. When applying a 5% discount, total revenue increased slightly to 6.18 million, although the average revenue per item decreased to around 728. This indicates that within the 0–5% range, the platform achieves high sales values while experiencing only a minor decline in average item value. This segment can be considered the most effective in balancing revenue generation against the cost of discounting.

As the discount level increases beyond 5%, a systematic decline is observed in both total revenue and average item value. At a 10% discount, total revenue drops to 5.73 million, with the average item value decreasing to approximately 691. At 15%, total revenue falls further to 5.32 million, and the average item value declines to around 640. A 20% discount results in a drop to 5.13 million in revenue and an average item value of about 615. The most unfavorable outcome appears at the 30% discount level, where total revenue falls to 4.35 million, and the average item value drops to roughly 522. In practice, this means that higher discount rates lead to significant reductions in both total revenue and transaction value.

From a business perspective, the findings suggest that customers do not increase their purchasing activity sufficiently to offset the revenue loss resulting from higher discounts. In the analyzed dataset, the number of sold items is relatively similar across discount tiers, reinforcing the conclusion that larger discounts do not meaningfully

increase transaction volume. Thus, aggressive price reductions are not an effective method for increasing sales in this case and instead result only in reduced order value and lower total revenue. This pattern is typical in markets where demand is relatively stable and driven more by consumer needs than by short-term price incentives.

A particularly interesting relationship emerges between the 0% and 5% levels. Although average item value is highest when no discount is applied, total revenue is marginally higher at the 5% discount level. This suggests that a small price reduction may act as an effective psychological incentive, encouraging some customers to complete a purchase, while the reduction remains small enough not to materially undermine profitability. From a discount-management standpoint, this indicates that moderate discounts, such as those in the 5–10% range, may be justified if they help acquire additional transactions or increase purchase frequency.

The analysis also leads to broader insights regarding the strategic role of discounting. In many organizations, discounts become a routine tool applied without regular evaluation of their impact. In the examined dataset, it is clear that discounts of 15%, 20%, or 30% do not generate added value from a revenue perspective. They reduce both the average transaction value and total revenue without delivering a meaningful increase in sales volume. As a result, such high discount levels should be used only in exceptional cases—for example, clearing end-of-line inventory, “clearance” campaigns, or activities aimed at rapid stock liquidation.

The analytical team concludes that the most rational approach for this dataset would be to focus the discount strategy on low-discount tiers, which demonstrate the most favorable balance between revenue and discount cost. Discounts in the 0–5% range appear optimal, as they produce the highest total revenue while maintaining relatively high average order values. Conversely, further increases in discount levels—especially to 20% or 30%—result in a clear deterioration of performance both at the unit level and in aggregate.

In summary, the analysis of discount policy effectiveness indicates that in this case, discounts should be treated as a precise tool, applied with careful consideration of their actual impact on revenue and average transaction value. High discounts not only fail to deliver the expected sales uplift but may even erode revenue and profitability. Therefore, it is recommended that the platform continue refining a strategy based on small, well-calculated discounts that support sales in a balanced manner without exerting excessive pressure on financial results.

Key Performance Indicators Analysis

Key performance indicators (KPIs) form the foundation for assessing the operational health of any organization operating in the online sales environment. While analyses of product categories, geographic regions, or discount levels provide detailed insight into specific areas of activity, it is the aggregation of these results at the KPI level that delivers a complete picture of platform performance. In the analyzed dataset, the most important indicators include total revenue, number of completed orders, average order value, and the average number of units purchased per order. Each of these metrics serves as an essential input for decision-making in sales management, offer optimization, and marketing budget allocation.

The analysis showed that during the examined period, the total sales value generated by the platform amounted to 32.87 million units. This result reflects high purchase activity among users and demonstrates the platform's operational scale. Considering the balanced distribution of sales across product categories and geographic regions, the total revenue level confirms that the platform operates in a stable manner and that revenue is evenly distributed across multiple market segments. Such a diversified revenue structure enhances organizational resilience to temporary declines in demand in specific segments or locations, which is a major advantage from a long-term strategic planning perspective.

Another key indicator is the number of orders, which totaled 50,000 during the analyzed period. This reflects a relatively large transactional base that allows for reliable statistical and predictive analyses. Additionally, the high order volume confirms that the platform maintains consistent customer engagement, which is particularly important given the competitive nature of the e-commerce market. A stable flow of orders is also crucial for maintaining operational continuity, both in logistics and customer service. The even distribution of orders over time reduces the risk of infrastructure overload during peak periods.

The average order value, which reached 657.33 units, suggests that the platform operates within the medium basket-value segment. The analytical team also noted that customers purchased on average two items per order, indicating that users typically buy more than one product per visit but do not engage in large-scale or bulk purchasing. This structure is characteristic of platforms offering a wide selection of products with varied price points. The average basket value suggests that customers are selecting products that fall neither into premium nor low-cost categories, which may indicate a well-balanced product offering and a strong positioning within the mid-range market segment.

The analysis of basket structure and average order value also provides insight into the effectiveness of cross-selling strategies and customer price sensitivity. In this dataset, the average number of items per basket suggests that customers make purchases either to

complement essential needs or in response to specific occasions. It does not reflect a "single-item purchase" profile typical for specialist platforms nor a bulk-purchase pattern characteristic of wholesale-oriented markets. Instead, the platform operates in a segment where users willingly add products to their baskets while maintaining a moderate approach toward expanding the scope of each transaction.

Average order value and basket structure are also significant in the context of discount policy analysis. The analytical team observed that the average order value decreases noticeably as discount levels increase. While low discounts do not significantly disrupt basket value, higher discounts reduce it to a degree that becomes unprofitable for the platform. In this context, the average basket value of 657 units can be interpreted as a natural equilibrium point between price and customer purchasing preferences. If the organization intends to increase the average basket value, it may need to implement initiatives aimed at boosting transaction-level value, such as product bundles, AI-driven recommendations, or loyalty programs that incentivize larger purchases.

The importance of KPIs extends beyond direct sales evaluation. KPIs also serve as reference points in financial forecasting and budgeting processes. They may form the basis for predictive models of demand and function as measures of marketing campaign effectiveness. For example, if new marketing activities lead to an increase in average order value or number of units per order, this may indicate that the campaign effectively enhances sales performance. Conversely, declining KPIs may serve as early warning signals indicating the need for strategic adjustments or optimization.

In summary, the analysis of key performance indicators confirms that the platform operates with stability in terms of both revenue and order volume. The average basket value and the number of purchased items suggest that customers demonstrate moderate yet steady purchasing activity. Combined with the balanced distribution of sales across categories and regions and predictable seasonality patterns, this indicates that the platform has a strong foundation for further scaling and for developing future sales and marketing strategies.

Business Conclusions

From a holistic perspective of the entire analysis, it can be concluded that the examined e-commerce platform operates in an environment characterized by a high degree of stability, both in terms of sales structure and demand dynamics. The normalized data model and the cross-sectional analyses provided insights into several complementary dimensions: product categories, geographic regions, seasonality, discount policy, and key performance indicators. The conclusions drawn from these areas form a coherent narrative indicating that the organization maintains a well-balanced product portfolio, diversified geographic risk, and predictable customer behavior—while also retaining headroom for further optimization.

In the area of product categories, the most notable feature is the strong balance between segments. Although the Beauty category reports the highest revenue, it does not dominate the market in extreme fashion, and other categories—Books, Fashion, Home & Kitchen, Electronics, and Sports—remain within a relatively narrow revenue range. This distribution reduces the risk associated with declining interest in a single category and reflects the platform's broad operating profile. At the same time, the slight advantage of Beauty and the consistently strong performance of Books indicate that these segments may serve as natural pillars for growth, especially if supported by well-designed marketing initiatives or loyalty programs.

The regional analysis leads to similar conclusions. Sales across the four examined regions—Middle East, North America, Asia, and Europe—are nearly equal, suggesting that the platform operates under a multi-regional model without significant disparities between local markets. Even though the Middle East generates the highest revenue, it does not significantly outpace the remaining regions, confirming an evenly distributed revenue base. From a business standpoint, this means that the organization is not overly dependent on any single geographic market, and each region contributes meaningfully to total sales. Such a structure supports financial stability and mitigates the impact of localized demand shocks on the broader business.

From a temporal perspective, the platform functions under conditions of moderate seasonality. Although distinct peaks appear—most notably the exceptionally strong performance in January 2023 and elevated sales during the summer months—the majority of monthly revenue figures fall within a narrow range. This indicates that the business is not driven primarily by a single peak period but instead relies on stable, repeatable demand throughout the year. On one hand, this is a positive operational signal, as it reduces the risk of dramatic revenue fluctuations; on the other, it highlights opportunities to better leverage weaker periods, such as February, through targeted promotional or engagement-oriented initiatives.

The most striking insights emerge from the discount-policy analysis. The data clearly shows that high discount levels do not deliver the expected revenue benefits. While the 0–5% range generates the highest total revenue with strong average item values, discounts in the 20–30% range lead to substantial declines in both total revenue and average transaction value. Combined with the finding that the number of sold items is similar across discount tiers, it becomes evident that increasing discounts does not yield a proportional increase in transaction volume. Economically, this signals that aggressive discounting does not improve sales effectiveness—instead, it reduces margins and may weaken customers' perception of product value.

The analysis of KPIs reinforces the picture of a stable, well-performing platform. Total revenue of 32.87 million, 50,000 orders, and an average basket value of 657 units with an average of two items per order suggest an operation with a lean yet healthy structure. Customers make purchases of moderate value, are not inclined toward large, one-off expenditures, yet use the platform consistently. This behavioral pattern can support the development of a stable revenue base, particularly if enhanced by initiatives that increase purchase frequency or basket size through cross-selling or recommendation mechanisms.

Combining the insights from all examined dimensions, it can be stated that the platform operates within a model that can be described as balanced and multidimensionally stable. This applies to product structure, geographic distribution, and temporal sales patterns alike. At the same time, the data reveals a clear area for optimization: the discount policy, which currently appears too broad and includes discount levels that do not provide economic benefits. In practice, this means that the organization is at a stage where it can focus not merely on building scale but on optimizing its operations—through deliberate shaping of the offer, pricing strategy, and customer-engagement methods.

These conclusions naturally lead to the recommendations presented in the next section of the report, which outline specific actionable directions for short- and medium-term implementation. These may include refinement of the discount strategy, deeper exploitation of natural seasonal peaks, strengthening high-potential categories, and advanced customer segmentation based on purchasing behavior. The subsequent part of the report translates these analytical insights into practical business actions.

Strategic Recommendations

Based on the analysis of sales data from 2022–2023, a number of strategic recommendations can be identified, the implementation of which could enhance the operational effectiveness and competitiveness of the platform. These recommendations are a direct consequence of the observed results and are intended to indicate actions that may strengthen both short-term performance and long-term growth potential. Given the platform's multidimensional stability and the balance visible in both product-category structure and regional revenue distribution, the recommendations focus primarily on areas where the analysis demonstrated meaningful opportunities for improvement or for leveraging observed trends more strategically.

The first area requiring particular attention is the discount policy. The analysis clearly demonstrated that the highest discount levels do not lead to a significant increase in transaction volume and instead reduce both average basket value and total revenue. These findings indicate that the platform should consider moving away from using discounts in the 15–30% range as a standard sales tool. Instead, it would be more effective to base promotional activity on low-discount tiers—particularly 0–5%—which exhibit the highest efficiency while preserving customer perception of product value. The platform may also consider implementing more precise discounting strategies based on customer segmentation, purchase history, or algorithmic price-recommendation models, enabling targeted sales increases where justified.

The second key area involves leveraging sales seasonality. Monthly results show several natural peaks that could serve as anchor points for targeted marketing efforts. The analytical team identified that sales in January, May, July, and August exceed those in other months, likely reflecting natural purchasing patterns related to the start of the year, the summer season, and back-to-school preparation periods. Meanwhile, February stands out as a relatively weaker month, representing an opportunity for focused promotional activities. The platform could introduce thematic campaigns, intensified marketing pushes, or additional purchasing incentives during periods of naturally lower customer activity. Such actions would help smooth seasonal fluctuations and increase average monthly revenue.

The third area highlighted by the analysis is product-category management. Although overall revenue distribution across categories is balanced, certain segments demonstrate slightly higher sales efficiency. The Beauty category emerged as the leading segment, suggesting high growth potential, particularly if supported by tailored marketing communication. The Books category also deserves attention: despite ongoing digitalization trends, it continues to generate substantial revenue, indicating either a loyal customer base or a strong product assortment. The platform could consider strengthening promotion in these segments while also exploring further opportunities in

high-performing areas such as Electronics and Home & Kitchen, which—despite their strong results—do not yet significantly outpace other categories.

From an operational-management perspective, the KPI analysis underscores the platform's stability and shows that the average basket value and number of purchased items remain steady. This indicates potential for initiatives aimed at increasing transaction value through cross-selling and upselling mechanisms. The platform could leverage historical purchase data and recommendation algorithms to offer complementary or alternative products, thereby increasing basket size. Such measures tend to be particularly effective on platforms with wide product assortments, where combining products across categories is natural for customers.

At the same time, the balanced regional revenue distribution suggests that the platform has substantial potential for further growth in existing markets. Since revenue contributions are evenly spread across regions, expansion efforts should not be limited to a single geographic area. This may involve continued localization of the product offering, enhancing presence in high-growth markets, and developing region-specific marketing initiatives. It may also be worthwhile to analyze which regions exhibit the strongest responsiveness to promotional campaigns, enabling more efficient allocation of resources in the future.

In summary, the presented recommendations aim to improve the platform's operational effectiveness and reinforce its competitive foundation. The organization has reached a stage where optimizing existing processes becomes more crucial than simply expanding scale. Actions focused on refining the discount strategy, better utilizing natural seasonality patterns, developing high-potential product categories, increasing basket value, and maintaining strong regional balance can collectively strengthen both revenue and profitability. The next section of the report presents final conclusions and potential directions for future research, which naturally extend the findings of this analysis.