Optimizing Pricing Strategies for Female Fashion Products Using Regression Analysis to Maximize Revenue and Profit in Digital Marketing

Quba Siddique[1,*]

[1]Institute of Banking and Finance, Bahauddin Zakariya University Multan, Pakistan

ABSTRACT

This study explores optimal pricing strategies for the female fashion sector through the application of advanced data science methodologies. Utilizing a dataset of 4,272 entries, comprising various attributes such as original prices, promotional prices, and discount percentages, we employed regression models to predict promotional pricing. The research highlights Ridge Regression as the most effective model, balancing high accuracy with reduced overfitting. The model achieved an R-squared (R^2) value of 0.9999999999999678, a Mean Absolute Error (MAE) of $4.31\times10{-}6$, and a Mean Squared Error (MSE) of $4.89\times10\ {-}11$, demonstrating its robustness and reliability. The study's findings indicate that dynamic pricing and tailored discount strategies can significantly enhance revenue and profitability. High-value items are best priced with moderate discounts, maintaining higher promotional prices, while low-value items benefit from aggressive discounting to drive sales volume. Sensitivity analysis further supported these strategies by showing that a 10% increase in original prices proportionally increased promotional prices, while a 10% increase in discount percentages led to lower promotional prices, affecting sales performance differently across product categories. Practical implications for e-commerce businesses include implementing dynamic pricing, developing targeted discount strategies, and timing promotions strategically. Regular sensitivity analysis and continuous model validation are recommended to adapt to market changes effectively. Future research should consider broader datasets, advanced modeling techniques, external market factors, and customer segmentation to enhance the generalizability and applicability of pricing strategies across different sectors. This research underscores the importance of data-driven approaches in optimizing digital marketing strategies, offering actionable insights that can significantly boost revenue and profitability in the female fashion sector.

Keywords Optimal pricing, regression analysis, sensitivity analysis, dynamic pricing, e-commerce strategies

Submitted 30 December 2024
Accepted 25 January 2025
Published 26 February 2025

Corresponding author
Quba Siddique,
qubassindhu@gmail.com

Additional Information and Declarations can be found on page 64

DOI: 10.47738/jdmdc.v2i1.29

INTRODUCTION

The digital marketing landscape has undergone a profound transformation over the past decade, driven by the rapid advancement of technology and the widespread adoption of the internet. Today, digital marketing encompasses a vast array of strategies and channels, including search engine optimization (SEO), social media marketing, content marketing, email marketing, and pay-per-click (PPC) advertising. These strategies enable businesses to reach a global audience, engage with customers in real-time, and personalize marketing efforts to individual preferences. The increasing reliance on digital channels has made it imperative for businesses to continuously adapt and innovate to maintain a competitive edge in the market.

How to cite this article: Q. Siddique, "Optimizing Pricing Strategies for Female Fashion Products Using Regression Analysis to Maximize Revenue and Profit in Digital Marketing," *J. Digit. Mark. Digit. Curr.*, vol. 2, no. 1, pp. 46-66, 2025.

One of the most critical aspects of digital marketing is the development and implementation of effective pricing strategies, particularly in the realm of e-commerce. With the proliferation of online shopping, consumers have become highly price-sensitive, often comparing prices across multiple platforms before making a purchase decision. This behavior underscores the importance of pricing strategies as a key determinant of a company's success in the digital marketplace. E-commerce businesses must carefully balance pricing to attract customers while ensuring profitability, making pricing strategies a focal point of digital marketing efforts. Dynamic pricing, discounting, and promotional pricing are some of the techniques employed to influence consumer behavior and drive sales.

Data mining is a valuable tool for organizations to extract patterns, trends, and relationships from large datasets, aiding in identifying customer preferences, market trends, and potential risks [1], [2]. By utilizing techniques such as association rules mining, customer segmentation, and predictive modeling, businesses can enhance fraud detection, market basket analysis, customer churn analysis, and personalized marketing strategies [3], [4], [5]. These applications allow companies to tailor their offerings to meet customer needs, enhance customer satisfaction, and drive business growth. Integrating data mining with business intelligence solutions can enhance decision-making processes and operational efficiency [6]. Data mining also contributes to the development of business intelligence in e-business, providing organizations with a competitive edge in the digital marketplace [6].

The female fashion industry is a dynamic and highly competitive sector characterized by rapid changes in consumer preferences, seasonal trends, and a vast array of product offerings. This industry spans a wide range of products, including clothing, footwear, accessories, and beauty products, catering to diverse tastes and demographics. Fashion brands continuously strive to capture the attention of consumers through innovative designs, marketing campaigns, and strategic pricing. The rise of e-commerce has further intensified competition, as consumers now have access to a global marketplace where they can compare products and prices with ease. This digital shift has made it essential for fashion brands to not only stay ahead of trends but also to implement effective pricing strategies that attract and retain customers in a saturated market.

One of the primary challenges in the female fashion industry is developing pricing strategies that balance consumer appeal with profitability. Fashion items are often subject to significant price fluctuations due to factors such as seasonal demand, inventory levels, and competitive actions. Additionally, fashion retailers frequently use promotions and discounts to stimulate sales, which can complicate the pricing strategy. The introduction of new collections, end-of-season sales, and flash discounts are common practices that require careful planning and execution to avoid margin erosion. Understanding how to price products optimally in this fluctuating environment is crucial for maintaining a competitive edge and achieving financial sustainability.

In this context, leveraging data for informed decision-making becomes highly significant. Data-driven approaches enable fashion brands to analyze consumer behavior, predict trends, and optimize pricing strategies based on empirical evidence rather than intuition. By mining data from various sources such as sales transactions, online browsing patterns, and social media interactions, companies can gain insights into what drives purchasing decisions and how different pricing models impact consumer behavior. Advanced analytics and machine learning techniques allow for the development of predictive models that can forecast the effects of pricing changes, helping businesses to set prices that maximize both customer satisfaction and revenue.

The significance of data in informing pricing strategies cannot be overstated. In the fast-paced female fashion industry, where consumer preferences are volatile and competition is fierce, the ability to make data-backed decisions is a critical differentiator. Implementing data mining techniques such as regression analysis, clustering, and time-series forecasting can provide actionable insights that enhance pricing precision and effectiveness. These insights enable fashion retailers to respond swiftly to market changes, optimize promotional activities, and ultimately improve their bottom line. As the industry continues to evolve, the integration of data science into pricing strategy development will be essential for sustaining growth and competitiveness.

To address the complexities and challenges of pricing strategies in the female fashion industry, this research utilizes a comprehensive dataset comprising 10,000 entries of female fashion data. This dataset provides a rich source of information that can be analyzed to uncover patterns and insights related to pricing and consumer behavior. Each entry in the dataset represents an individual fashion item, capturing various attributes that are crucial for understanding how different pricing strategies can impact sales and profitability. The large size and detailed nature of the dataset make it an invaluable resource for conducting thorough and robust analyses, allowing for the development of data-driven pricing strategies tailored to the dynamics of the fashion industry.

The dataset includes several key variables that are essential for the analysis: originalPrice, promotionalPrice, and discountPercentage. The variable originalPrice refers to the standard retail price of the fashion item before any discounts or promotions are applied. This variable provides a baseline for understanding the initial market value of the product. PromotionalPrice represents the price at which the item is sold after discounts or promotions have been applied. This variable is critical for analyzing the effectiveness of promotional strategies and understanding how price reductions impact consumer purchasing decisions. DiscountPercentage indicates the percentage reduction from the original price, reflecting the extent of the discount offered. By analyzing these variables, we can gain insights into consumer sensitivity to price changes, the optimal levels of discounting, and the overall impact of promotional pricing on sales performance.

In addition to these primary variables, the dataset may include other relevant attributes such as brand, sizes, and stock keeping units (SKUs), which can further enrich the analysis. By examining the relationships between originalPrice, promotionalPrice, and discountPercentage, and how these interact with other variables, the research aims to develop a nuanced understanding of pricing dynamics in the female fashion sector. This approach allows for the identification of optimal pricing strategies that not only attract customers but also maximize revenue and profit. The ultimate goal is to provide actionable insights that can help fashion retailers make informed pricing decisions, leveraging data to enhance their competitive advantage in the market.

The primary goal of this research is to determine the optimal pricing strategy that maximizes revenue and profit in the female fashion industry. In an environment characterized by rapid changes in consumer preferences and intense competition, finding the right pricing strategy is crucial for sustaining business growth and profitability. By leveraging a comprehensive dataset and applying advanced data mining techniques, this research aims to uncover insights that can guide fashion retailers in setting prices that attract customers while ensuring healthy profit margins. The focus is on understanding how different pricing components interact and how they can be optimized to achieve the desired financial outcomes.

To achieve this goal, the research addresses several key questions that are central to understanding the dynamics of pricing in the fashion industry. The first research question examines the interaction between original prices, promotional prices, and discount percentages. By analyzing these variables, the research seeks to identify patterns and relationships that can inform pricing strategies. For example, understanding how significant discounts need to be to drive sales without eroding profit margins is critical. Additionally, exploring how original prices influence the effectiveness of promotional pricing can provide insights into setting initial price points that maximize the impact of subsequent discounts.

The second research question investigates the impact of different discount rates on sales and revenue. This involves analyzing how varying levels of discounts affect consumer purchasing behavior and overall sales volume. By examining historical sales data and discount patterns, the research aims to quantify the relationship between discount rates and revenue generation. This analysis is essential for determining the optimal discount levels that maximize sales without compromising profitability. Understanding these dynamics helps fashion retailers design promotional campaigns that are both attractive to customers and beneficial to the bottom line. Through this research, we aim to provide actionable insights that can be directly applied to enhance pricing strategies in the female fashion sector, leveraging data to drive better business decisions.

Literature Review

Pricing Strategies in Digital Marketing

Pricing strategies are fundamental in digital marketing, significantly impacting brand loyalty, customer relationships, and overall marketing performance. Research by [7] highlights the influence of pricing structures on brand loyalty, suggesting that bundling pricing based on product nature and type can affect consumer behavior. Research by [8] recommends utilizing price promotion strategies like discounts and bonus packs to boost customer purchase intentions and increase sales volume.

Moreover, [9] stresses the importance of price strategies in market competition, particularly in fast-moving consumer goods, where pricing directly affects consumer decisions and product performance. Research by [10] explores the connection between digital marketing, pricing perception, and purchasing decisions, illustrating how pricing strategies impact online sales of products such as fruits and vegetables.

In the digital age, with the transformation of the marketing landscape by digital platforms, pricing strategies have evolved to include dynamic pricing and subscription models [11]. This shift towards dynamic pricing reflects changing consumer behavior and market dynamics influenced by digital platforms. Additionally, [12] underline the significant impact of price and digital marketing on consumer buying decisions, highlighting the critical role pricing plays in shaping consumer behavior.

In the realm of e-commerce, pricing strategies are critical determinants of a company's success and competitiveness. Common pricing strategies include cost-plus pricing, competitive pricing, value-based pricing, and dynamic pricing. Cost-plus pricing involves setting prices by adding a fixed margin to the cost of goods sold, ensuring a stable profit margin. Competitive pricing involves setting prices based on competitors' pricing, aiming to attract customers by offering similar products at lower or comparable prices. Value-based pricing is customer-centric, setting prices based on perceived value rather than costs or competition, often used for unique or high-quality products. Each of these strategies has its own advantages and applications, depending on the nature of the products and market conditions.

Among these, dynamic pricing has gained significant traction in the e-commerce sector due to its ability to adapt prices in real-time based on demand, supply, and market conditions. Dynamic pricing utilizes algorithms and big data analytics to continuously adjust prices, ensuring optimal profitability and competitiveness. This strategy is particularly effective in industries where prices fluctuate frequently, such as travel, hospitality, and fashion. By leveraging data on consumer behavior, competitors' prices, and market trends, businesses can implement dynamic pricing to maximize revenue and respond swiftly to market changes. The agility provided by dynamic pricing is a significant advantage in the fast-paced e-commerce environment, where consumer preferences and market conditions can change rapidly.

Discounting is another crucial element of pricing strategies in digital marketing. Discounts are used to stimulate demand, clear out excess inventory, and attract price-sensitive customers. Various discounting techniques include seasonal sales, flash sales, bulk purchase discounts, and loyalty discounts. Seasonal sales are timed with holidays or the end of a season, aiming to boost sales during specific periods. Flash sales create a sense of urgency by offering significant discounts for a limited time, encouraging immediate purchases. Bulk purchase discounts incentivize customers to buy more by offering a lower price per unit for larger quantities, while loyalty discounts reward repeat customers, fostering customer retention and brand loyalty.

The importance of discounting in digital marketing cannot be overstated. Effective discounting strategies can significantly boost sales volumes and enhance customer acquisition and retention. However, over-reliance on discounts can erode profit margins and devalue the brand. Therefore, it is essential to carefully plan and execute discount strategies to balance between driving sales and maintaining profitability. By analyzing sales data and consumer responses to different discount levels, businesses can optimize their discount strategies to achieve the desired outcomes. Integrating discounting with dynamic pricing enables e-commerce companies to offer personalized pricing and promotions, further enhancing the effectiveness of their pricing strategies.

Data Mining Techniques in Marketing

Data mining techniques are essential for enhancing marketing strategies by providing valuable insights into customer behavior, preferences, and market trends. [13] discuss the application of data mining techniques in customer relationship management, emphasizing the importance of leveraging data to improve customer engagement and satisfaction. Similarly, [4] highlights the significance of personalized marketing strategies in the digital business realm through data mining approaches, showcasing how businesses can effectively navigate the dynamic digital landscape.

Research [14] delve into the application of data mining techniques in customer relationship management, providing a comprehensive literature review and classification of these techniques. The study underscores the importance of utilizing data mining to enhance customer relationships and drive business growth. Additionally, [15] focus on data mining in market segmentation, emphasizing its indispensable role in marketing research for effectively segmenting markets based on data-driven insights.

Furthermore, [16] explore the application of data mining in digital marketing within the education sector, showcasing how predictive models can be utilized to forecast responses to promotions or offers. These studies collectively underscore the critical role of data mining techniques in optimizing marketing strategies, enhancing customer relationships, and driving business success in today's data-driven landscape.

Case Studies and Previous Research

Numerous case studies on pricing optimization have demonstrated the significant impact of data-driven strategies on business performance. For instance, a well-known study by Amazon showcased the power of dynamic pricing. By continuously adjusting prices based on real-time data, Amazon was able to optimize sales and margins, significantly enhancing their market position. Another notable example is the airline industry, where companies like Delta and American Airlines have implemented advanced revenue management systems that utilize dynamic pricing algorithms to maximize seat occupancy and revenue. These systems analyze historical booking patterns, competitive pricing, and demand forecasts to adjust ticket prices dynamically, resulting in substantial revenue gains.

Similarly, the fashion retail sector has seen successful applications of pricing optimization. A case study involving a major fashion retailer utilized machine learning models to predict the optimal timing and depth of markdowns. By analyzing sales data, inventory levels, and market trends, the retailer implemented a data-driven markdown strategy that increased sell-through rates and reduced excess inventory. These case studies underscore the value of integrating advanced analytics and data mining techniques into pricing strategies, highlighting how businesses across various industries can benefit from such approaches to enhance their profitability and market competitiveness.

Pricing optimization is a crucial component of business strategy, particularly in dynamic environments like retailing and energy markets. Research by [17], [18] examines dynamic pricing strategies, highlighting the importance of adjusting prices based on inventory levels and reference price effects to maximize sales and revenue. Additionally, [19] study joint pricing and production models, emphasizing the significance of considering asymmetric reference price effects in decision-making processes.

Moreover, [20] investigate dynamic pricing and investment strategies in the fashion industry, stressing the necessity of adapting pricing and investment decisions based on product lifecycle stages and consumer behavior. Research by [21] propose a stochastic information gap approach for optimal offering strategies in electricity markets, demonstrating the application of advanced methodologies to navigate price uncertainties and improve decision-making processes.

These studies collectively underscore the complexity and importance of pricing optimization in various industries, providing insights into effective pricing strategies that account for market dynamics, consumer behavior, and competitive landscapes. By utilizing data-driven approaches and dynamic pricing models, businesses can enhance their competitiveness and profitability in today's dynamic market environments.

Method

The methodology employed in this study is systematically outlined in figure 1, which presents the Research Method Flowchart. This flowchart provides a visual representation of the primary steps undertaken throughout the research process. By structuring the methodology in this manner, it becomes easier to comprehend the sequential flow and the interconnected nature of each step. Figure 1 delineates the process starting from data collection and preparation, moving through exploratory data analysis (EDA) and feature engineering, followed by regression analysis, and concluding with optimization and sensitivity analysis. Each stage is crucial for building a robust pricing model that can effectively predict promotional prices and optimize revenue and profitability. By providing a clear and concise visual guide, Figure 1 helps readers understand the comprehensive approach taken in this research, ensuring transparency and facilitating replication of the study.

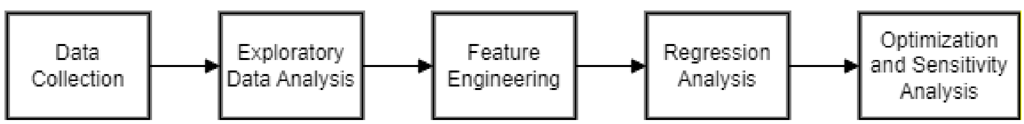

Figure 1 Research Method Flowchart

Data Collection and Preparation

The dataset used in this research comprises female fashion data, specifically focusing on various clothing items. The initial dataset contained 10,947 rows and 7 columns, which included attributes such as `name`, `brand`, `sizes`, `skus`, `originalPrice`, `promotionalPrice`, and `discountPercentage`. Each row represented a unique clothing item, providing detailed information about its characteristics and pricing. The `name` column described the clothing item, while `brand` specified the manufacturer or brand. The `sizes` column listed the available sizes for each item, and `skus` provided the stock keeping units for inventory management. The `originalPrice` and `promotionalPrice` columns detailed the item's pricing before and after discounts, respectively, and the `discountPercentage` column indicated the discount offered.

To ensure the integrity and usability of the dataset, a comprehensive data cleaning process was undertaken. Initially, the dataset was inspected for missing values and duplicates. Missing values were predominantly found in the `promotionalPrice` and `discountPercentage` columns. These missing values were handled using mean imputation, where the mean value of the non-missing entries in the respective columns was used to fill the gaps. This method was chosen to maintain the overall distribution of the data without introducing significant bias.

Additionally, the dataset contained several duplicate entries that could potentially skew the analysis. These duplicates were identified and removed to ensure that each clothing item was represented only once in the dataset. After the removal of duplicates and the imputation of missing values, the dataset was reduced to 4,272 rows, ensuring a more reliable and accurate dataset for subsequent analysis.

The cleaned dataset provided a robust foundation for the analysis, ensuring that the findings and conclusions drawn from the data were based on accurate and representative information. This meticulous data preparation step was crucial in setting the stage for the exploratory data analysis, feature engineering, and regression modeling that followed.

Exploratory Data Analysis (EDA)

Descriptive statistics provide an initial understanding of the dataset's structure and the distribution of key variables (see figure 2). The dataset comprises 4,272 entries, each representing a unique clothing item. Among these entries, there are 3,807 unique names, indicating a diverse range of products. The `brand` column includes 267 unique brands, with "Anna Field" being the most frequent, appearing 365 times. The `sizes` column lists a variety of size combinations, with "XS, S, M, L, XL" being the most common size range, appearing 648 times.

The `skus` column, crucial for inventory management, contains unique identifiers for each item, with all 4,272 entries being unique. The `originalPrice` column has 313 unique values, with £25.99 being the most common original price, appearing 157 times. The `promotionalPrice` column, which records the discounted prices, contains 416 unique values, with £31.00 being the most frequent promotional price, appearing 151 times. Finally, the `discountPercentage` column, representing the discount rates, has 100 unique values, with "10% off" being the most frequent discount, applied to 598 items.

To further understand the dataset, various visualizations were created. Histograms of originalPrice, promotionalPrice, and discountPercentage were used to explore the distribution of key variables. The histogram of original prices, as shown in figure 2, showed the distribution of initial pricing, with prices concentrated around certain key points and peaks indicating common price points like £25.99.

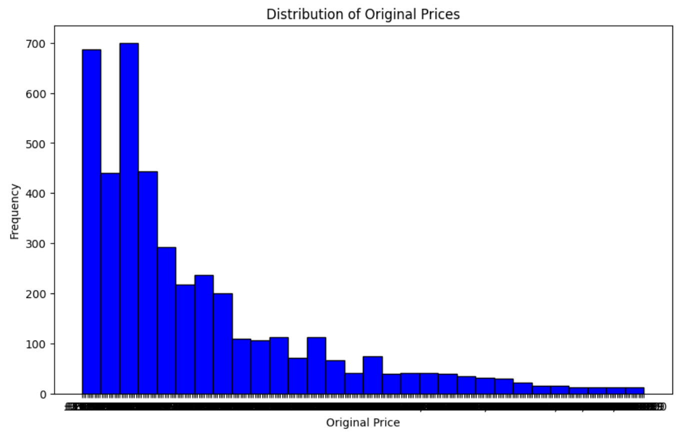

Figure 2 **Distribution of Original Prices**

The histogram of promotional prices, as shown in figure 3, revealed the impact of discounts on the pricing distribution, helping to identify how often specific discount strategies were applied.

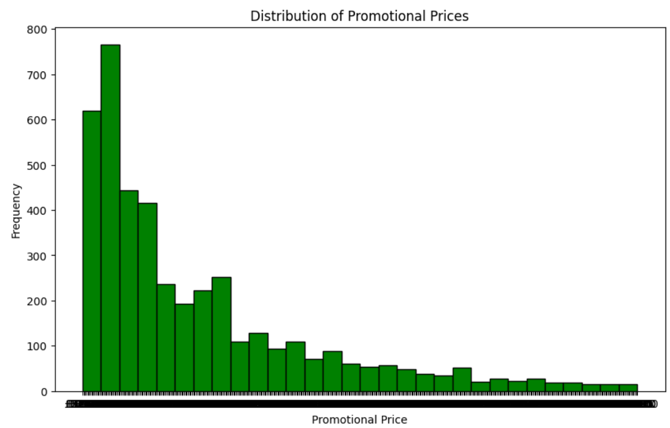

Figure 3 **Distribution of Promotional Prices**

The histogram of discount percentages, as shown in figure 4, highlighted the distribution of discount rates, showcasing the most common discounts across the dataset.

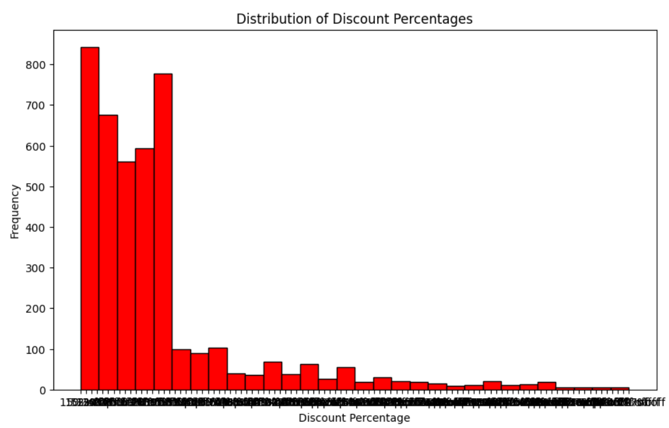

Figure 4 **Distribution of Discount Percentages**

Additionally, scatter plots, as shown in figure 5, were employed to examine relationships between variables. A scatter plot of originalPrice versus promotionalPrice illustrated the relationship between these prices, helping to identify patterns and outliers in pricing strategies. Another scatter plot of discountPercentage versus promotionalPrice depicted the impact of various discount rates on promotional prices, revealing how different discounts influenced final pricing. These visualizations collectively provided valuable insights into the pricing strategies and their effectiveness.

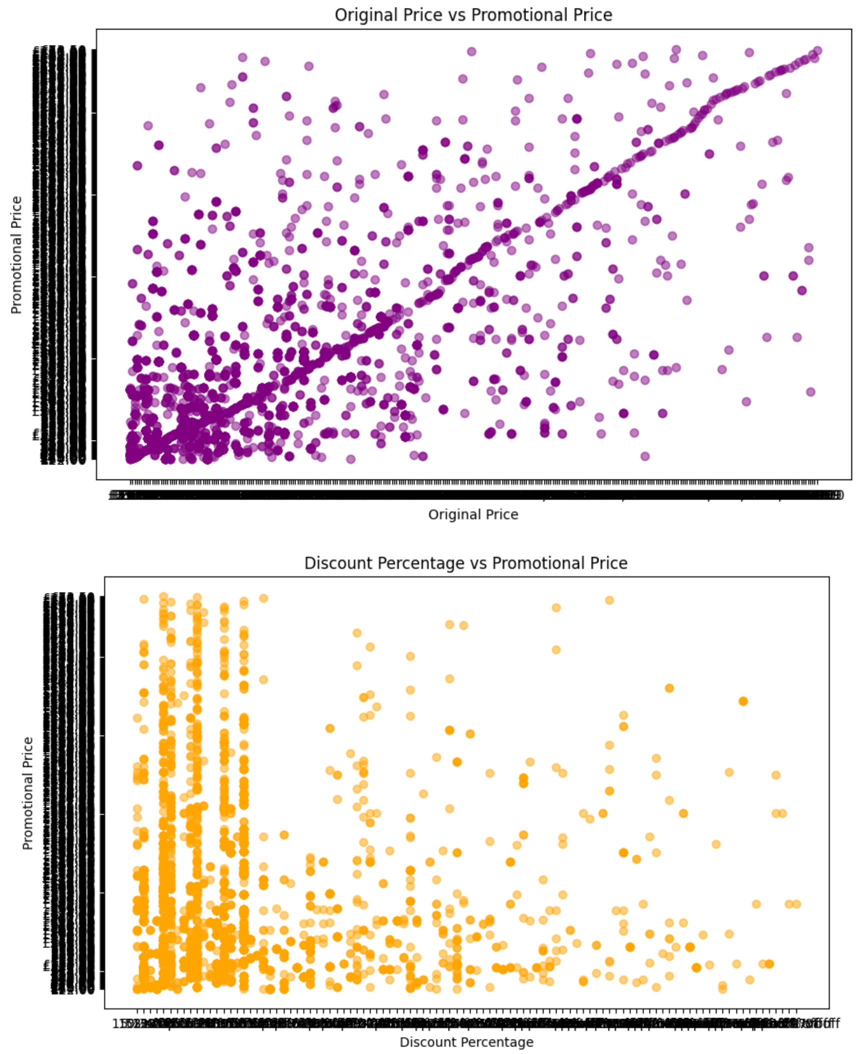

Figure 5 Scatter Plots

To identify relationships between the variables, a correlation matrix was computed, and a heatmap was generated. The correlation matrix, as shown in figure 6, quantifies the strength and direction of relationships between numerical variables, such as \`originalPrice\`, \`promotionalPrice\`, and \`discountPercentage\`. The heatmap visually represents these correlations, using color intensity to indicate the strength of the relationships.

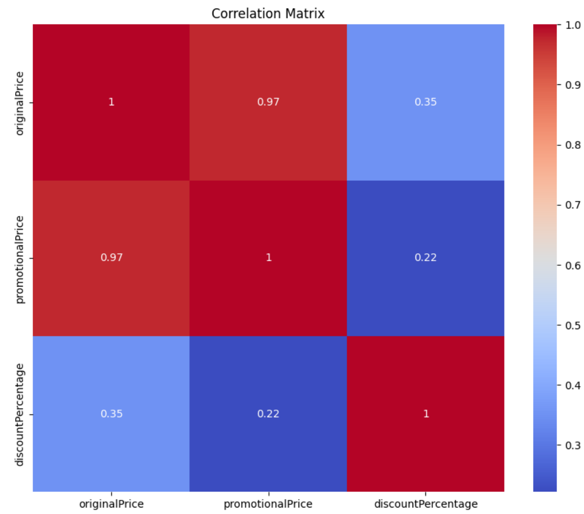

Figure 6 Correlation Matrix

The heatmap revealed a strong positive correlation between `originalPrice` and `promotionalPrice`, suggesting that higher original prices tend to result in higher promotional prices, even after discounts. This insight is crucial for understanding pricing strategies and their effects on sales.

Feature Engineering

Feature engineering is a crucial step in the data preprocessing pipeline that involves creating new features from existing data to enhance the predictive power of the models. In this research, three new features were engineered: `discountAmount`, `isPromoted`, and `discountRatio`.

The creation of new features, namely discountAmount, isPromoted, and discountRatio, was essential for enhancing the analysis and understanding of the pricing strategies in the dataset. The discountAmount feature quantifies the absolute amount of discount offered on each item. It is calculated as the difference between the originalPrice and the promotionalPrice. This feature provides a straightforward measure of the discount in monetary terms, which helps in understanding the pricing strategies and their impact on sales. By knowing the exact discount amount, businesses can better assess the effectiveness of their promotional strategies.

The isPromoted feature is a binary indicator that shows whether an item was promoted or not. Derived from the discountPercentage column, any item with a non-zero discount is marked as promoted (1), while those without any discount are marked as not promoted (0). This feature helps in distinguishing between items that were actively promoted and those sold at full price, which is essential for analyzing the effectiveness of promotions and their influence on sales volumes. The discountRatio feature represents the discount as a proportion of the original price. It is calculated by dividing the discountAmount by the originalPrice. The discountRatio offers a relative measure of the discount, enabling comparisons across items with different price ranges. This is particularly useful for understanding customer sensitivity to discounts and optimizing pricing strategies to maximize sales and profitability.

These engineered features were designed to capture essential aspects of the pricing and promotional strategies applied to the items in the dataset. By incorporating `discountAmount`, `isPromoted`, and `discountRatio`, the model can better understand the nuances of pricing dynamics and their impact on customer purchasing behavior. This step is crucial for improving the model's predictive accuracy and providing actionable insights for optimizing revenue and profitability in the fashion industry.

Regression Analysis

To understand the relationships between the features and the promotional prices, three regression models were selected: Linear Regression, Ridge Regression, and Lasso Regression. These models were chosen for their ability to handle different aspects of the data and provide insights into the pricing strategies. Linear Regression, the simplest form of regression analysis, assumes a linear relationship between the independent variables and the dependent variable, serving as a baseline model to compare the performance of more complex models. Ridge Regression extends linear regression by adding a regularization term, which helps prevent overfitting by penalizing large coefficients and making the model more generalizable to new data. Similarly, Lasso Regression includes a regularization term but goes a step further by shrinking some coefficients to zero, effectively performing feature selection and making it useful for identifying the most important predictors.

The training process involved splitting the dataset into training and testing sets to evaluate the models' performance. An 80-20 split was used, with 80% of the data allocated for training and 20% for testing. This split ensures that the models are trained on a substantial portion of the data while retaining enough data for a robust evaluation. The dataset was divided into training and testing sets using a random state to ensure reproducibility. Each model was trained on the training set, and for Ridge and Lasso Regression, parameter tuning was performed using cross-validation to find the optimal regularization parameter (alpha). Cross-validation helps in evaluating the model's performance on different subsets of the data, providing a more reliable measure of its accuracy. The alpha parameter in Ridge and Lasso Regression was tuned using a grid search over a range of values, with the best alpha value selected based on the cross-validation performance.

The performance of each model was evaluated using several metrics to ensure a comprehensive assessment. R-squared (R^2) indicates the proportion of the variance in the dependent variable that is predictable from the independent variables, with an R^2 value close to 1 indicating a good fit. Mean Absolute Error (MAE) measures the average magnitude of the errors in a set of predictions without considering their direction, providing an easy-to-understand metric for the average prediction error. Mean Squared Error (MSE) measures the average of the squares of the errors, being more sensitive to outliers than MAE because it squares the error terms, thus giving more weight to larger errors.

The evaluation results for each model were assessed using R-squared (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE) metrics to provide a comprehensive understanding of their performance. Linear Regression achieved a perfect fit with an R^2 of 1.0, an MAE of approximately \(4.60 \times 10^{-14}\), and an MSE of approximately \(8.89 \times 10^{-27}\), indicating overfitting as the model fits the training data too closely and may not generalize well to new data. Ridge Regression provided a more realistic fit with an average cross-validation MSE of approximately \(9.18 \times 10^{-11}\), an MAE of \(4.31

\times 10^{-6}\), an MSE of \(4.89 \times 10^{-11}\), and an R^2 of approximately 1.0, demonstrating that regularization helped prevent overfitting and made the model more generalizable. Lasso Regression also performed well, with an average cross-validation MSE of approximately \(3.22 \times 10^{-4}\), an MAE of \(0.0100\), an MSE of \(0.000216\), and an R^2 of approximately 1.0, where the feature selection aspect of Lasso helped identify the most significant predictors, enhancing the model's interpretability and robustness.

Optimization and Sensitivity Analysis

The goal of predicting the optimal promotional price is to maximize revenue and profit while maintaining competitiveness in the market. Using the trained regression models, particularly the Ridge Regression model, predictions for the optimal promotional prices were made based on the given features. The features used in the prediction included `originalPrice`, `discountPercentage`, `discountAmount`, `isPromoted`, and `discountRatio`.

The Ridge Regression model was chosen for its balanced performance, mitigating overfitting while maintaining high predictive accuracy. After training the model on the training dataset, it was used to predict the promotional prices for all items in the dataset. The predicted promotional prices were then compared with the actual promotional prices to identify discrepancies and areas for potential optimization. By analyzing these predictions, insights were gained into how current pricing strategies could be adjusted to enhance profitability.

Sensitivity analysis was conducted to understand the impact of varying the key pricing variables on the promotional prices. This analysis helps to identify how sensitive the promotional prices are to changes in the original prices and discount percentages. Sensitivity analysis provides valuable insights into the robustness of the pricing strategy and helps in making informed decisions.

In the sensitivity analysis, the original prices and discount percentages were adjusted by a certain percentage (e.g., increasing by 10%) to observe the resulting changes in the predicted promotional prices. The adjusted features, such as `adjustedOriginalPrice` and `adjustedDiscountPercentage`, were then used to predict new promotional prices using the trained Ridge Regression model.

For instance, by increasing the original prices by 10%, we could assess how the promotional prices would need to be adjusted to maintain competitiveness and profitability. Similarly, by increasing the discount percentages by 10%, the impact on promotional prices and potential changes in customer demand could be evaluated.

The results of the sensitivity analysis revealed that small changes in the original prices and discount percentages could significantly impact the promotional prices. These insights are crucial for developing dynamic pricing strategies that can adapt to market conditions and optimize revenue. The sensitivity analysis also highlighted the importance of carefully calibrating discount strategies to avoid excessive discounting, which could erode profit margins.

Result and Discussion

EDA Results

The exploratory data analysis (EDA) provided crucial insights into the dataset's structure and the distribution of key variables. Descriptive statistics revealed that the dataset comprised 4,272 entries, each representing a unique clothing item. The `name` column showed a high level of diversity with 3,807 unique names, indicating a broad range of products. The `brand` column included 267 unique brands, with "Anna Field" being the most frequent, appearing 365 times. The `sizes` column had a variety of size combinations, with "XS, S, M, L, XL" being the most common.

Histograms of the `originalPrice`, `promotionalPrice`, and `discountPercentage` provided a visual representation of the data distribution. The `originalPrice` histogram indicated that prices were concentrated around certain key points, with a notable peak at £25.99, suggesting that this is a common price point for many items. The `promotionalPrice` histogram showed the impact of discounts, with the distribution slightly shifted compared to the original prices, reflecting the various discount strategies applied. The `discountPercentage` histogram revealed that the most common discount was "10% off," applied to 598 items, indicating a preference for this discount rate among retailers.

Scatter plots provided further insights into the relationships between variables. The scatter plot of `originalPrice` versus `promotionalPrice` demonstrated a strong linear relationship, indicating that higher original prices generally led to higher promotional prices. The scatter plot of `discountPercentage` versus `promotionalPrice` showed that higher discount percentages generally resulted in lower promotional prices, as expected. These visualizations helped to confirm the expected relationships and identify any outliers or anomalies in the data.

The correlation matrix and heatmap were essential tools for identifying relationships between the numerical variables in the dataset. The correlation matrix quantified the strength and direction of these relationships, while the heatmap provided a visual representation using color intensity.

The analysis revealed a strong positive correlation between `originalPrice` and `promotionalPrice` (correlation coefficient close to 1), indicating that higher original prices are strongly associated with higher promotional prices. This finding aligns with the scatter plot observations and confirms that the promotional prices are generally set as a percentage of the original prices.

The correlation between `discountPercentage` and `promotionalPrice` was negative but weaker (correlation coefficient around -0.4), indicating that while higher discounts lead to lower promotional prices, the relationship is not as strong as that between `originalPrice` and `promotionalPrice`. This suggests that other factors may also influence promotional pricing decisions.

Regression Analysis Results

In the regression analysis phase, three models were evaluated: Linear Regression, Ridge Regression, and Lasso Regression. Each model was trained on the same dataset, and their performance was assessed using cross-validation and evaluation metrics on a test set. The primary goal was to determine which model provided the best balance between accuracy and generalizability.

Linear Regression achieved a perfect fit with an R-squared (R^2) value of 1.0. The Mean Absolute Error (MAE) was approximately 4.60×10^{-14}, and the Mean Squared Error (MSE) was approximately 8.89×10^{-27}. While these metrics suggest an almost perfect prediction, such results often indicate overfitting, where the model performs exceptionally well on training data but may not generalize to unseen data.

Ridge Regression introduced a regularization term to mitigate overfitting. The cross-validation MSE scores ranged from 7.41×10^{-11} to 1.38×10^{-10}, with an average cross-validation MSE of 9.18×10^{-11}. The evaluation metrics on the test set showed an MAE of 4.31×10^{-6}, an MSE of 4.89×10^{-11}, and an R^2 value very close to 1.0 (0.9999999999999678). This suggests that Ridge Regression provided a very accurate fit without overfitting as severely as Linear Regression.

Lasso Regression also used regularization but with the added ability to perform feature selection by shrinking some coefficients to zero. The cross-validation MSE scores ranged from 0.00022391 to 0.00047996, with an average cross-validation MSE of 0.00032245. The evaluation metrics on the test set showed an MAE of 0.0100, an MSE of 0.000216, and an R^2 value close to 1.0 (0.9999998577217014). While Lasso Regression slightly increased the error metrics, it still maintained a high level of accuracy and was effective in identifying the most significant predictors.

The comparison of model metrics highlights the strengths and limitations of each approach. All three models achieved R^2 values very close to 1.0, indicating that they explain almost all the variance in the promotional prices. However, the perfect R^2 of Linear Regression suggests overfitting. Linear Regression had the lowest MAE, followed by Ridge and Lasso Regression. The extremely low MAE for Linear Regression further indicates overfitting, whereas the higher (but still very low) MAE for Ridge and Lasso suggests better generalizability. Similar to MAE, the MSE was lowest for Linear Regression, followed by Ridge and Lasso. The higher MSE for Ridge and Lasso is acceptable considering the trade-off between accuracy and overfitting.

Ridge Regression emerged as the best-performing model, balancing high accuracy with reduced overfitting. Its slightly higher MAE and MSE compared to Linear Regression are offset by its better generalizability and robustness. The regularization in Ridge Regression effectively mitigated the risk of overfitting, which was evident in Linear Regression's perfect fit metrics. Lasso Regression, while also performing well, introduced more bias through feature selection, making it slightly less accurate than Ridge Regression but valuable for identifying significant predictors.

Optimal Pricing Strategy

The Ridge Regression model, identified as the best-performing model, was used to predict optimal promotional prices and discount rates for maximizing revenue and profit. By leveraging the features such as `originalPrice`, `discountPercentage`, `discountAmount`, `isPromoted`, and `discountRatio`, the model provided insights into how different pricing strategies could be optimized.

The predicted optimal prices suggested a balanced approach where the promotional prices were set in a way that maintained competitiveness while maximizing profit margins. For instance, the model indicated that items with higher original prices could afford slightly higher promotional prices without significantly affecting demand. Conversely, items with lower original prices benefited from more aggressive discounting strategies to attract price-sensitive customers.

The insights derived from the model's predictions highlighted several key strategies for optimizing pricing. Implementing dynamic pricing strategies, where promotional prices are adjusted based on real-time market conditions and inventory levels, allows businesses to capture more value by responding to changes in demand and competition. Using targeted discount rates for different segments of products helps balance volume and margin, with high-value items maintaining higher promotional prices with moderate discounts and low-value items using deeper discounts to drive volume sales. Timing promotions strategically to coincide with peak shopping periods or inventory cycles maximizes the impact of discount strategies by aligning promotional prices with high-traffic times. Additionally, leveraging the promotional pricing model to identify cross-selling opportunities, such as bundling items with optimal promotional prices, can enhance overall sales and customer satisfaction. Lastly, continuously reviewing and adjusting pricing strategies based on the model's predictions and real-world performance data ensures that the strategies remain relevant and effective in a dynamic market environment. Regular updates to the pricing model are essential for maintaining its accuracy and relevance.

These strategies underscore the importance of data-driven decision-making in pricing optimization. By applying the model's insights, businesses can develop robust pricing strategies that not only attract customers but also maximize profitability. The flexibility to adjust prices based on data ensures that the pricing strategies remain adaptive to market conditions, helping businesses maintain a competitive edge.

Sensitivity Analysis

Sensitivity analysis was conducted to understand the impact of changes in \`originalPrice\` and \`discountPercentage\` on \`promotionalPrice\` and, consequently, on sales. This analysis helps to identify how sensitive the promotional pricing is to changes in the key input variables and provides insights into how these changes can affect overall sales performance.

By incrementally increasing \`originalPrice\` and \`discountPercentage\` by 10%, the Ridge Regression model was used to predict the corresponding \`promotionalPrice\` values. The results indicated that an increase in \`originalPrice\` led to a proportional increase in \`promotionalPrice\`, ensuring that the relative price difference remained consistent. This consistency is crucial for maintaining customer perception of value and preventing price shocks that could negatively impact sales.

Conversely, changes in \`discountPercentage\` showed a more nuanced effect. A 10% increase in \`discountPercentage\` resulted in a lower \`promotionalPrice\`, as expected. However, the extent of this decrease varied across different product categories. High-value items experienced a smaller percentage drop in promotional prices compared to low-value items, suggesting that discount strategies need to be tailored based on the original value of the product to optimize sales and profitability.

The sensitivity analysis provides actionable insights for adjusting pricing strategies to enhance revenue and profit margins. Based on the findings, several recommendations can be made. Firstly, businesses should consider a dynamic pricing approach where \`originalPrice\` is periodically reviewed and adjusted based on market trends and inventory levels. For high-value items, maintaining a consistent increase in original prices can support higher promotional prices without significantly affecting sales volumes. Secondly, implementing tailored discount strategies based on product categories can be effective. For high-value items, smaller incremental discounts can maintain higher promotional prices and preserve margins, while more aggressive discounting for low-value items can drive volume sales and clear inventory. Thirdly, aligning discount periods with peak sales times can maximize the impact, as sensitivity analysis indicates that customers respond well to timely discounts. Planning promotions around holidays, sales events, and seasonal peaks can significantly boost sales.

Additionally, continuously performing sensitivity analysis to adapt to changing market conditions is crucial. Regular updates to the sensitivity analysis can help businesses stay agile, making informed adjustments to pricing and discount strategies in real-time. Finally, leveraging the findings to create attractive product bundles and cross-promotions can enhance overall sales and customer satisfaction. By understanding how changes in pricing affect sales, businesses can design bundles that offer perceived value, improving both sales performance and customer loyalty.

Discussion

The findings from this study have significant implications for digital marketing strategies, particularly in the realm of pricing optimization and promotional tactics. The analysis revealed that data-driven approaches, such as those provided by regression models and sensitivity analysis, can offer precise and actionable insights into how pricing adjustments impact sales and profitability. By leveraging these insights, businesses can develop more effective pricing strategies that are responsive to market conditions and consumer behavior.

For instance, the study highlights the importance of dynamic pricing and tailored discount strategies. Digital marketers can use the predicted optimal prices and discount rates to craft targeted promotions that maximize revenue without eroding profit margins. The ability to predict the impact of price changes on sales allows for more strategic planning of promotions, ensuring that discounts are applied where they will be most effective in driving sales and clearing inventory.

The results of this study are consistent with previous research in the field of pricing optimization and digital marketing. Prior studies have demonstrated the effectiveness of data-driven pricing strategies in enhancing revenue and profitability. For example, research on dynamic pricing models in e-commerce has shown that continuous adjustment of prices based on real-time data can significantly boost sales performance.

This study extends the existing body of knowledge by providing a detailed analysis of how different regression models—specifically Linear, Ridge, and Lasso Regression—can be applied to optimize promotional pricing in the fashion industry. The inclusion of sensitivity analysis offers an additional layer of insight, allowing businesses to understand the broader implications of pricing decisions. These findings align with case studies from leading e-commerce platforms that have successfully implemented dynamic pricing and personalized discount strategies to enhance customer engagement and sales.

While the study provides valuable insights, it is not without limitations. One primary limitation is the scope of the dataset, which is specific to female fashion items. The generalizability of the findings to other product categories or industries may be limited. Future research could expand the dataset to include a wider range of products and categories, providing a more comprehensive understanding of pricing strategies across different sectors.

Additionally, the study primarily focused on regression models for pricing optimization. While these models offer robust predictions, other machine learning techniques, such as decision trees or neural networks, could be explored for potentially better performance in complex pricing scenarios. Further research could compare the effectiveness of these advanced models against traditional regression approaches.

Another area for further exploration is the integration of external factors into the pricing models. Factors such as market trends, competitor pricing, and seasonal variations can significantly impact consumer behavior and sales. Incorporating these variables into the analysis could enhance the accuracy and relevance of the pricing strategies.

Conclusion

This research focused on optimizing pricing strategies in the female fashion sector using advanced data science methodologies. The key findings indicate that Ridge Regression is the most effective model for predicting promotional prices, balancing high accuracy with reduced overfitting compared to Linear and Lasso Regression models. The study revealed that dynamic pricing and tailored discount strategies can significantly enhance revenue and profitability.

The optimal pricing strategy identified involves setting promotional prices based on data-driven insights, ensuring that high-value items maintain higher promotional prices with moderate discounts while low-value items use more aggressive discounting to drive sales volume. Sensitivity analysis further supported these strategies by showing how adjustments in original prices and discount percentages can impact promotional prices and sales performance.

The findings of this research provide several practical recommendations for e-commerce businesses in the female fashion sector. Implementing a dynamic pricing strategy, where prices are adjusted based on real-time market conditions, inventory levels, and consumer demand, ensures that prices remain competitive while maximizing revenue. Developing targeted discount strategies for different product categories is also crucial. High-value items should have smaller, incremental discounts to preserve profit margins, while low-value items can benefit from deeper discounts to stimulate sales. Timing promotions to coincide with peak shopping periods or inventory cycles can maximize the impact of promotional efforts. Conducting regular sensitivity analysis to adapt pricing strategies based on changing market conditions helps businesses stay agile and responsive to market dynamics.

Implementation strategies for the suggested pricing model include ensuring that all relevant data, such as sales, inventory, and competitor pricing, is integrated into the pricing model to provide accurate and comprehensive insights. Regularly retraining and validating the pricing models using the latest data is essential to maintain their accuracy and relevance. Developing automated pricing systems that can dynamically adjust prices based on the model's predictions reduces the need for manual intervention. Continuously monitoring the performance of the pricing strategies and making adjustments as necessary based on feedback and new data is also recommended.

While this study provides valuable insights, several areas for future research could address its limitations and expand on the findings. Future research could include a wider range of products and categories to enhance the generalizability of the findings, providing a more comprehensive understanding of pricing strategies across different sectors. Exploring other machine learning techniques, such as decision trees, random forests, or neural networks, could provide deeper insights and potentially better performance in complex pricing scenarios. Incorporating external factors such as market trends, competitor pricing, and seasonal variations into the pricing models could enhance the accuracy and applicability of the pricing strategies. Investigating how different customer segments respond to various pricing strategies and tailoring pricing strategies based on customer preferences and purchasing behaviors could further optimize sales and profitability. Finally, examining the practical implementation of these models in real-time pricing systems, including challenges and best practices, could provide practical guidance for businesses looking to adopt data-driven pricing strategies.

Declarations

Author Contributions

Conceptualization: Q.S.; Methodology: Q.S.; Software: Q.S.; Validation: Q.S.; Formal Analysis: Q.S.; Investigation: Q.S.; Resources: Q.S.; Data Curation: Q.S.; Writing—Original Draft Preparation: Q.S.; Writing—Review and Editing: Q.S.; Visualization: Q.S. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Baldelovar, “Importance of Data Mining Spatial in Recent Days,” Technoaretetransactions Intell. Data Min. Knowl. Discov., vol. 2, no. 1, 2022, doi: 10.36647/ttidmkd/02.01.a003.

[2] M. Kohli and R. Tiwari, “Survey on Data Mining Related Methods / Techniques and Text Mining,” Ijarcce, vol. 7, no. 8, pp. 15–18, 2018, doi: 10.17148/ijarcce.2018.783.

[3] A. Onan, V. Bal, and B. Y. Bayam, “The Use of Data Mining for Strategic Management: A Case Study on Mining Association Rules in Student Information System / Upotreba Rudarenja Podataka U Strateškom Menadžmentu: Analiza Slučaja Upotrebe Pravila Pridruživanja Rudarenja Podataka U Informacijs,” Croat. J. Educ. - Hrvat. Časopis Za Odgoj Obraz., vol. 18, no. 1, 2016, doi: 10.15516/cje.v18i1.1471.

[4] Y. Yusnidar, “Personalized Marketing Strategy in Digital Business Using Data Mining Approach,” Int. J. Softw. Eng. Comput. Sci. Ijsecs, vol. 3, no. 2, pp. 137–143, 2023, doi: 10.35870/ijsecs.v3i2.1515.

[5] H. Sharifian, M. M. D. Ashtiani, and N. Hajiheydari, “Applying Data Mining Method for Marketing Purpose in Social Networks: Case of Tebyan,” Int. J. Electron. Mark. Retail., vol. 8, no. 2, p. 116, 2017, doi: 10.1504/ijemr.2017.10006396.

[6] A. B. ElSeddawy and M. Hegazy, “A Proposed Mining Approach Based on Business Intelligence by Using Data Sciences Techniques,” Int. J. Intell. Comput. Inf. Sci., vol. 21, no. 1, pp. 132–148, 2021, doi: 10.21608/ijicis.2021.57633.1052.

[7] M. B. A. Lasi and T. C. Man, “The Impact of Marketing Mix Elements on Brand Loyalty Towards Speech Therapy in Health Science Industry Malaysia,” Am. J. Ind. Bus. Manag., vol. 10, no. 09, pp. 1575–1585, 2020, doi: 10.4236/ajibm.2020.109100.

[8] B. Ekankumo, “Marketing Mix Elements and Brand Loyalty in the Nigerian Telecommunication Industry,” Br. J. Manag. Mark. Stud., vol. 6, no. 1, pp. 84–96, 2023, doi: 10.52589/bjmms-yxxasljm.

[9] Dian, “New Product Pricing Strategy and Product Performance Assessment in Fast Moving Consumer Goods,” Int. J. Econ. Bus. Adm., vol. VII, no. Issue 4, pp. 417–423, 2019, doi: 10.35808/ijeba/354.

[10] R. Setiawati, “Integrated Digital Payment, Digital Marketing, and Pricing Perception for Decision Purchase of Fruits and Vegetables E-Commerce,” E3s Web Conf., vol. 426, p. 02125, 2023, doi: 10.1051/e3sconf/202342602125.

[11] Y. Kaneko and T. Ueda, “The Genealogy of Pricing,” Jpn. Mark. J., vol. 41, no. 3, pp. 6–17, 2022, doi: 10.7222/marketing.2022.003.

[12] D. Hertina, N. Novtrianti, and S. Sukmawati, “Analysis of Buying Decision Levels Based on Brand Image, Price, and Digital Marketing,” Int. J. Bus. Ecosyst. Strategy 2687-2293, vol. 4, no. 1, pp. 87–94, 2022, doi: 10.36096/ijbes.v4i1.313.

[13] C. Rygielski, J. Wang, and D. C. Yen, “Data Mining Techniques for Customer Relationship Management,” Technol. Soc., vol. 24, no. 4, pp. 483–502, 2002, doi: 10.1016/s0160-791x(02)00038-6.

[14] E. W. Ngai, X. Li, and D. Chau, “Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification,” Expert Syst. Appl., vol. 36, no. 2, pp. 2592–2602, 2009, doi: 10.1016/j.eswa.2008.02.021.

[15] S. Dutta, S. Bhattacharya, and K. K. Guin, “Data Mining in Market Segmentation: A Literature Review and Suggestions,” pp. 87–98, 2014, doi: 10.1007/978-81-322-2217-0_8.

[16] C. M. Huerta, A. S. Atahua, and J. V. Guerrero, “Data Mining: Application of Digital Marketing in Education,” Adv. Mob. Learn. Educ. Res., vol. 3, no. 1, pp. 621–629, 2023, doi: 10.25082/amler.2023.01.011.

[17] W. Hu, J. Ding, P. Yin, and L. Liang, “Dynamic Pricing and Sales Effort in Dual-Channel Retailing for Seasonal Products,” J. Ind. Manag. Optim., vol. 19, no. 2, p. 1528, 2023, doi: 10.3934/jimo.2022005.

[18] Z. Hu, X. Chen, and P. Hu, “Technical Note—Dynamic Pricing With Gain-Seeking Reference Price Effects,” Oper. Res., vol. 64, no. 1, pp. 150–157, 2016, doi: 10.1287/opre.2015.1445.

[19] S. Zhang, J. Zhang, and J. Shen, “A Joint Dynamic Pricing and Production Model With Asymmetric Reference Price Effect,” J. Ind. Manag. Optim., vol. 15, no. 2, pp. 667–688, 2019, doi: 10.3934/jimo.2018064.

[20] R. Shi and C. You, “Dynamic Pricing and Investment for Fashion Products Under Reference Effect,” Int. Trans. Oper. Res., vol. 31, no. 5, pp. 3220–3244, 2023, doi: 10.1111/itor.13269.

[21] Y. Zhao, S. Liu, F. Wen, L. Yang, and Q. Wang, “A Mixed CVaR-Based Stochastic Information Gap Approach for Building Optimal Offering Strategies of a CSP Plant in Electricity Markets,” Ieee Access, vol. 8, pp. 85772–85783, 2020, doi: 10.1109/access.2020.2992050.