Optimize inventory management using prescriptive analytics.

By: Somasekhar Sai Gutti, Thushar Mohan Gaddigopula

University of Colorado Denver

Course Title: Prescriptive Analytics with Analytics

Professor: Yu Du

Table of Contents

Project Overview

Introduction:

The goal of the Summer Fashion Products Analysis project is to investigate and evaluate a dataset that includes details on different summer fashion items that are sold on an internet marketplace. The purpose of this research is to learn more about the performance, trends, and preferences of summer fashion items. Given the seasonal fluctuations and dynamic nature of the fashion sector, it might be beneficial for both buyers and sellers to have a thorough awareness of the features and demand for summer products.

Context:

Businesses need to be aware of industry trends and consumer behavior in order to make well-informed decisions, especially with the growing importance of e-commerce platforms (Balakrishnan, 2021). The dataset utilized for this project includes information on product characteristics, costs, ratings, and sales figures, giving users a thorough understanding of the summer fashion items that are offered on the platform. The project's analysis of this data attempts to pinpoint popular patterns, evaluate the influence of different aspects on product success, and provide sellers with insights that might inform their inventory and marketing plans.

Numerous factors will be examined in the analysis, such as product categories, pricing schemes, customer feedback, and the impact of particular attributes on sales. Through an exploration of these factors, the initiative aims to offer practical insights to buyers seeking to understand summer fashion preferences as well as merchants seeking to maximize their products.

This initiative adds to our understanding of how seasonal variations affect customer decisions in the apparel sector while also being in line with current e-commerce trends. Businesses can use the results of this analysis to increase customer satisfaction, expand their product offers, and maintain their competitiveness in the ever-changing world of online shopping (Jallouli, 2019).

In conclusion, the goal of the Summer Fashion Products Analysis project is to identify trends in consumer behavior and product performance in the summer fashion market. This will allow the project to provide useful information to stakeholders in the fashion retail industry.

Objectives:

The following are the main goals of the Summer Fashion Products Analysis project:

- **Determine Popular Summer Fashion Trends**: In order to determine the most well-liked and in-demand summer fashion items, the project will conduct a thorough analysis of the dataset. In order to identify the dominant market trends, this entails taking into account variables like sales volume, customer ratings, and other pertinent information.
- **Assess Pricing Strategies:** One of the most important parts of the analysis is looking at the connection between sales success and product pricing. Finding sensible pricing schemes for summer fashion items that suit consumer tastes and market conditions is the aim.
- **Evaluate Customer Satisfaction:** Examining ratings and reviews is the main way to gauge customer satisfaction, which is a crucial project metric. This goal seeks to identify the elements that influence either positive or negative feedback in order to offer suggestions for enhancing the general customer experience.
- **Examine Product Attributes:** In the context of summer fashion, the project will look into how different product attributes, such as color, size, and style, affect sales and consumer preferences. Comprehending these inclinations is imperative in customizing product offerings to satisfy customer demands.
- **Recognize Shipping Dynamics:** Examining the impact of various shipping options on consumer decisions and product success in the summer is a crucial component of this analysis. The purpose of this goal is to shed light on the importance of prompt and practical delivery options (Kaabi, 2019).
- **Find High-Performing Vendors:** The initiative will identify vendors who regularly produce summer fashion items that exhibit strong performance. This goal is to assist other sellers in optimizing their methods to increase their market presence by showcasing effective tactics and procedures.
- **Discover Seasonal Urgency**: This study will look at the existence and efficacy of urgency banners in promoting sales, particularly in the summer. Understanding how these marketing strategies affect consumer behavior and purchase decisions is the goal of this objective.

- **Publish Actionable Insights:** The project's final objective is to present actionable insights and a summary of the major discoveries. In the summer fashion e-commerce market, the analysis will empower sellers and customers by simplifying complex data into actionable recommendations.
- **Contribute to Industry Knowledge:** In addition to achieving particular goals, the project aims to advance knowledge about consumer behavior and market dynamics in the fashion industry, with a special emphasis on the distinctive features of the summertime.

Dataset Description

Data Source

The retail.csv dataset was obtained from an extensive retail inventory that included a range of product listings. It records various aspects of retail operations, such as pricing, product data, sales performance, and client feedback.

Data Structure

The dataset contains several key columns which provide valuable insights into the product listings and sales patterns. Important columns include:

- Title: Product name.
- Price: Selling price.
- Retail_Price: Manufacturer's suggested retail price.
- Currency_Buyer: Transaction currency.
- Units_Sold: Total sales volume for each product.
- Uses_Ad_Boosts: Indicator of advertisement boost usage.
- Rating: Average product rating.
- Rating_Count: Total number of ratings.
- Merchant_Rating: Average rating of the seller.
- Product_Id: Unique identifier for products.

Exploratory Data Analysis (EDA)

Descriptive Statistics

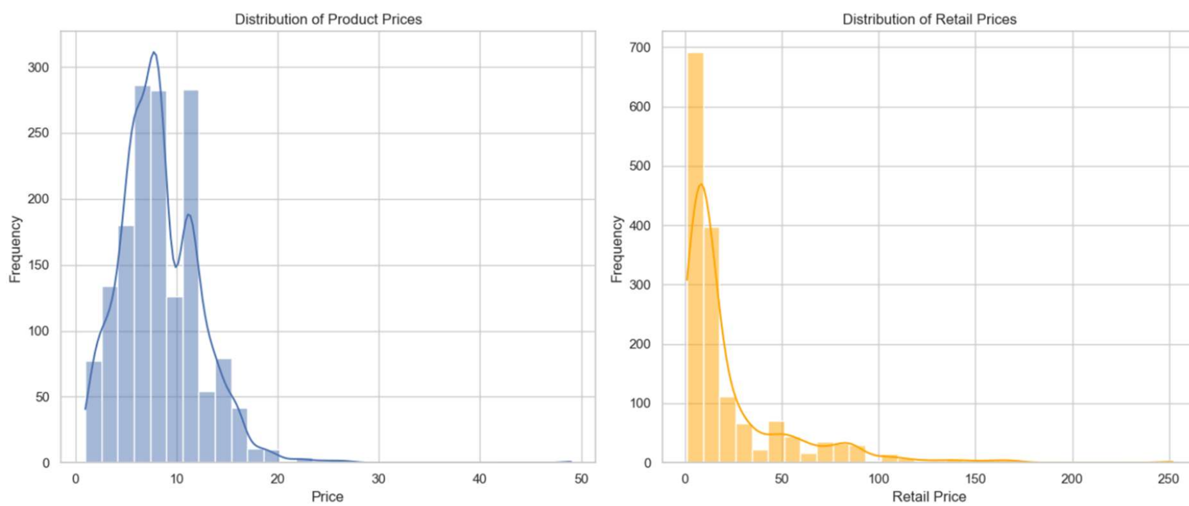
The descriptive statistics highlight important features of the summer fashion products under analysis and offer a thorough overview of the dataset's key columns.

- **Price:**
  The dataset's products have an average price of $8.33 and a standard deviation of $3.93. The observed price ranges from a minimum of $1.00 to a maximum of $49.00. This range illustrates how different summer fashion items have different pricing strategies.
- **Retail Price:**
  When it comes to retail prices, the standard deviation is $30.36 and the mean is $23.29. The retail prices vary greatly; they start at $1.00 and go up to $252.00 at the highest. This broad range indicates that the products have different market positioning and approaches.
- **Sold Units:**
  4,339 units were sold on average, with a noteworthy standard deviation of 9,357. Although the minimum quantity sold is one, suggesting products with low demand, some products have sold up to a significant maximum of 100,000 units. The variance in product demand within the dataset is highlighted by this.
- **Rating:**
  In terms of product ratings, the standard deviation is 0.52 and the mean is 3.82. The observed rating ranges from a minimum of 1.00 to a maximum of 5.00. These ratings serve as an indicator of customer satisfaction and shed light on how the summer fashion items were received in general.
- **Number of ratings:**
  890 is the average rating count, and 1,984 is the standard deviation. The count ranges from 0 at the lowest to 20,744 at the highest, highlighting the wide variation in ratings that each product has obtained. This measure sheds light on the products' level of engagement and popularity.

- **Merchant Evaluation:**
  With a standard deviation of 0.20, merchants are rated 4.03 on average. A merchant's rating can be as low as 2.33 or as high as 5.00. The overall evaluation of the dependability and credibility of sellers in the summer fashion e-commerce market is influenced by these merchant ratings.
  In conclusion, these descriptive statistics provide a quantitative basis for comprehending the distribution and variability within the dataset's important columns, offering insightful knowledge about the workings of the summer fashion industry.

Data Visualization

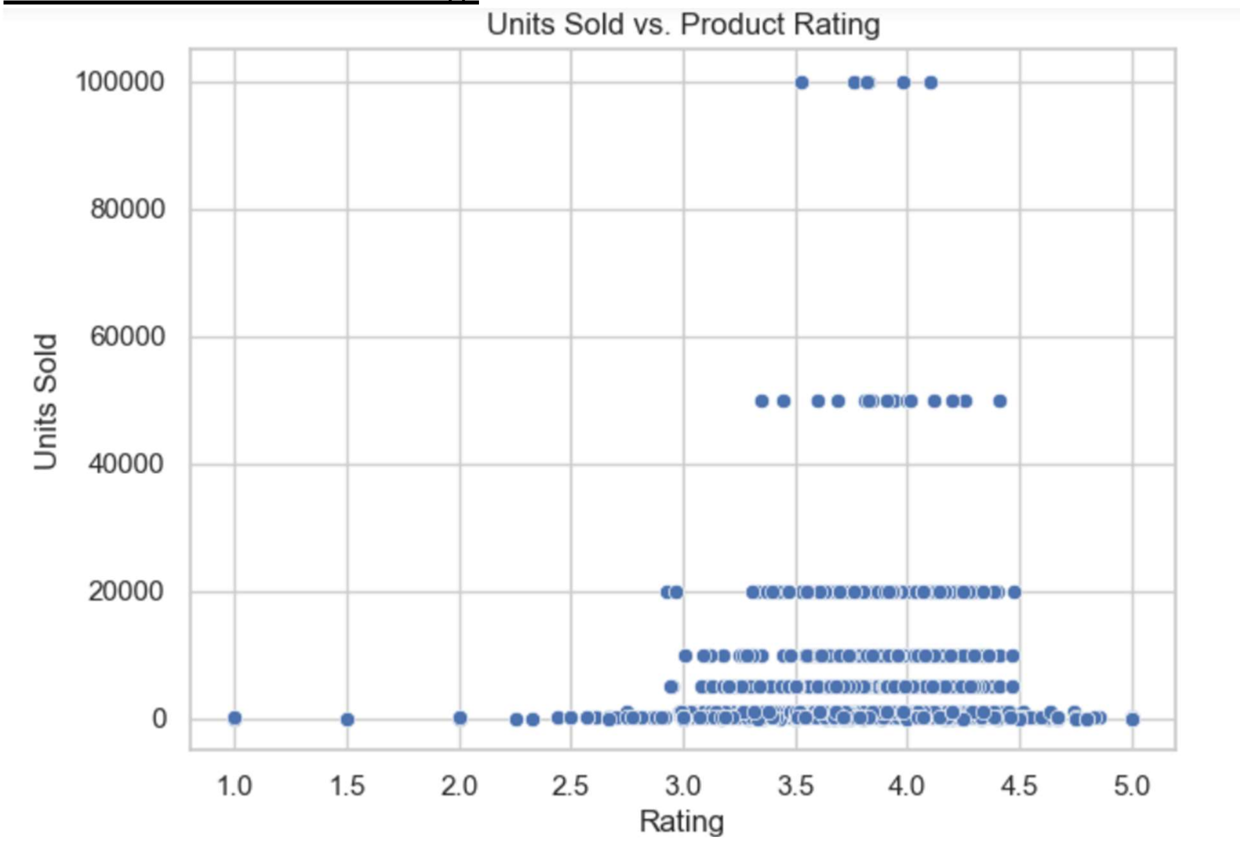
Distribution of Product Prices and Retail Prices:



The distribution of product prices is displayed in the first plot. There is a concentration of products in the lower price range, as evidenced by the majority of products having prices under \$20, peaking at \$8.

The retail price distribution is depicted in the second plot. Like the selling prices, a large proportion of the products had retail pricing under \$30, with a peak of about \$10. This implies that the products are often priced between lower and mid-range.

Units Sold vs Product Rating:



The correlation between product ratings and units sold is represented visually by the scatter plot. There doesn't seem to be a significant linear relationship between sales volume and ratings. A broad variety of sales volumes are displayed by products with ratings between 3.5 and 4.5, suggesting that variables other than product ratings may affect sales.

Methodology

Data Cleaning:

Missing Values

Missing values in columns like has_urgency_banner, urgency_text, and merchant_profile_picture have been replaced with 'Unknown'. For columns with fewer missing values (product_color, product_variation_size_id, origin_country), missing entries were imputed with the most frequent value.

Outlier Handling

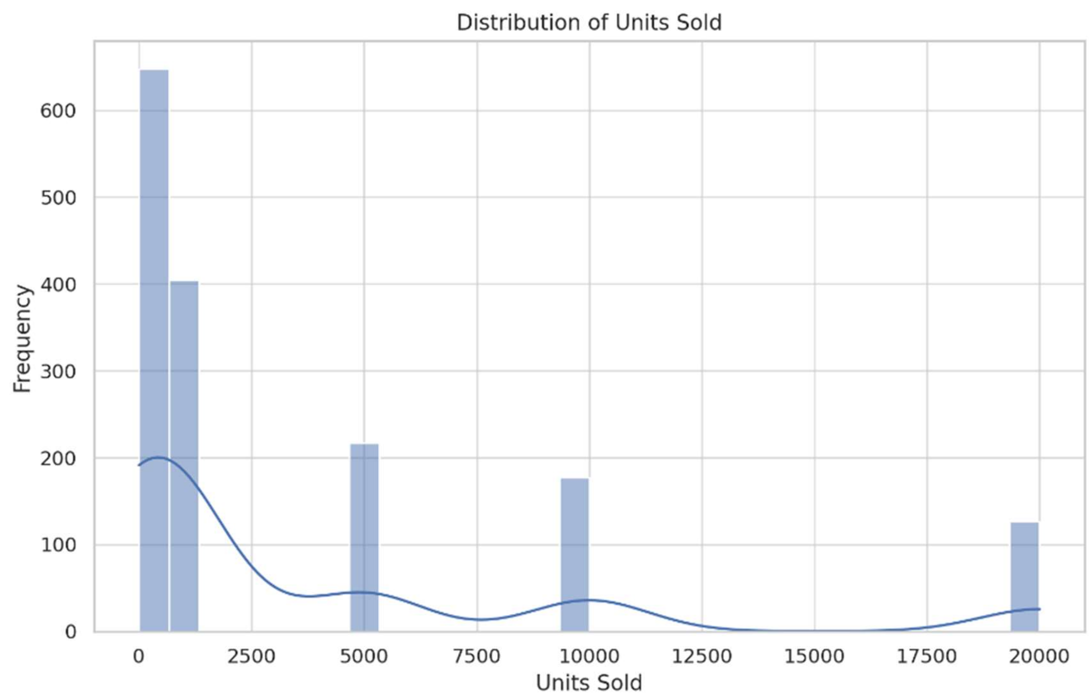
Outliers in the units_sold column were capped at the 95th percentile. This reduces the impact of extreme values on subsequent analyses.

Most columns have no missing values after these cleaning procedures. The values over the 95th percentile cap are no longer present in the units_sold column.

Now that the dataset is more uniform and cleaner, it is ready for the next phase of our investigation, which entails developing models for inventory optimization and demand forecasting as well as exploratory data analysis (EDA).

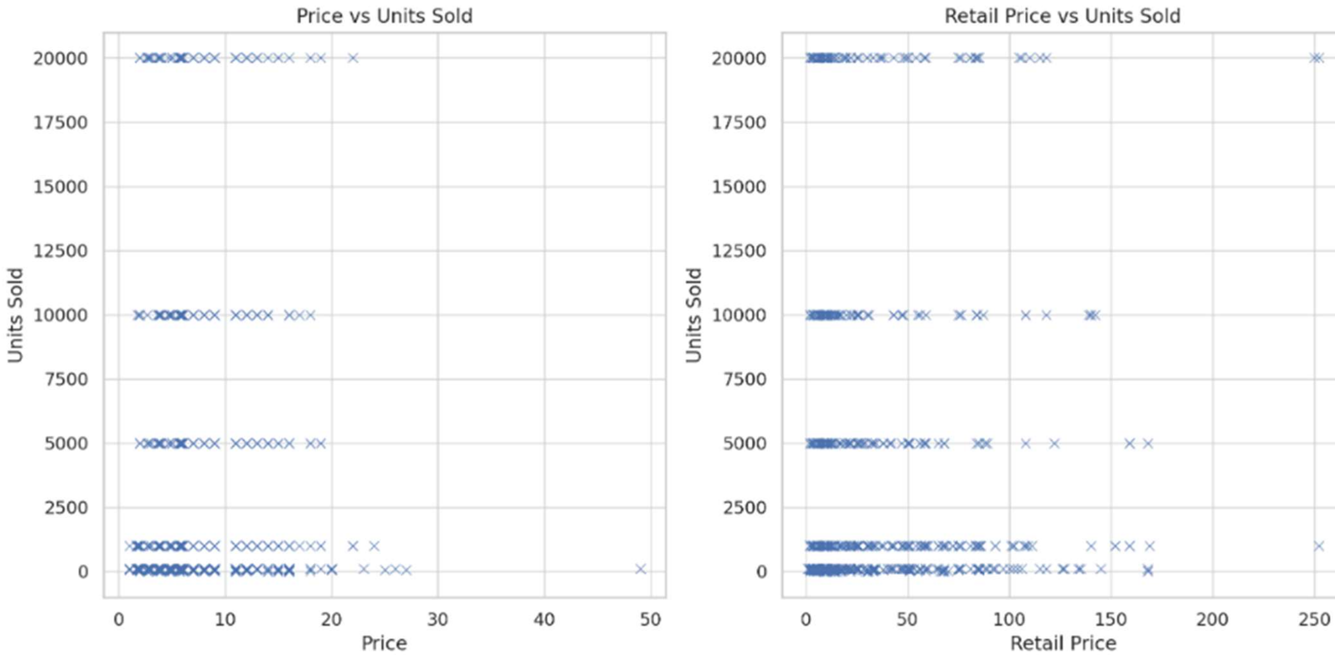
Data Analysis Techniques

Sales Trend Analysis:



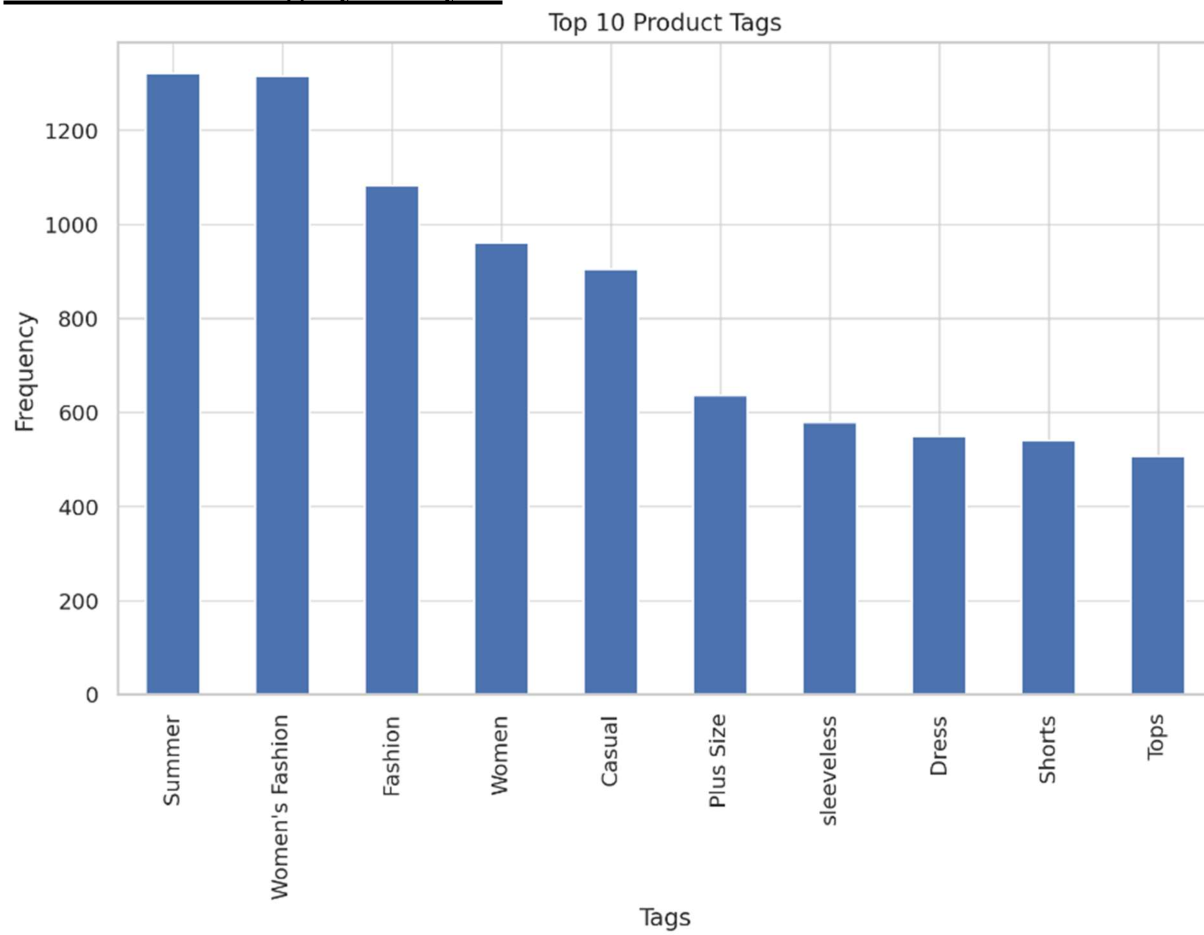
A concentration of products with lower sales volumes is visible in the distribution of units sold, which displays a wide range. This implies that some products may be top sellers, but a large percentage of products have low sales figures (Kaabi, 2019).

Price Analysis:



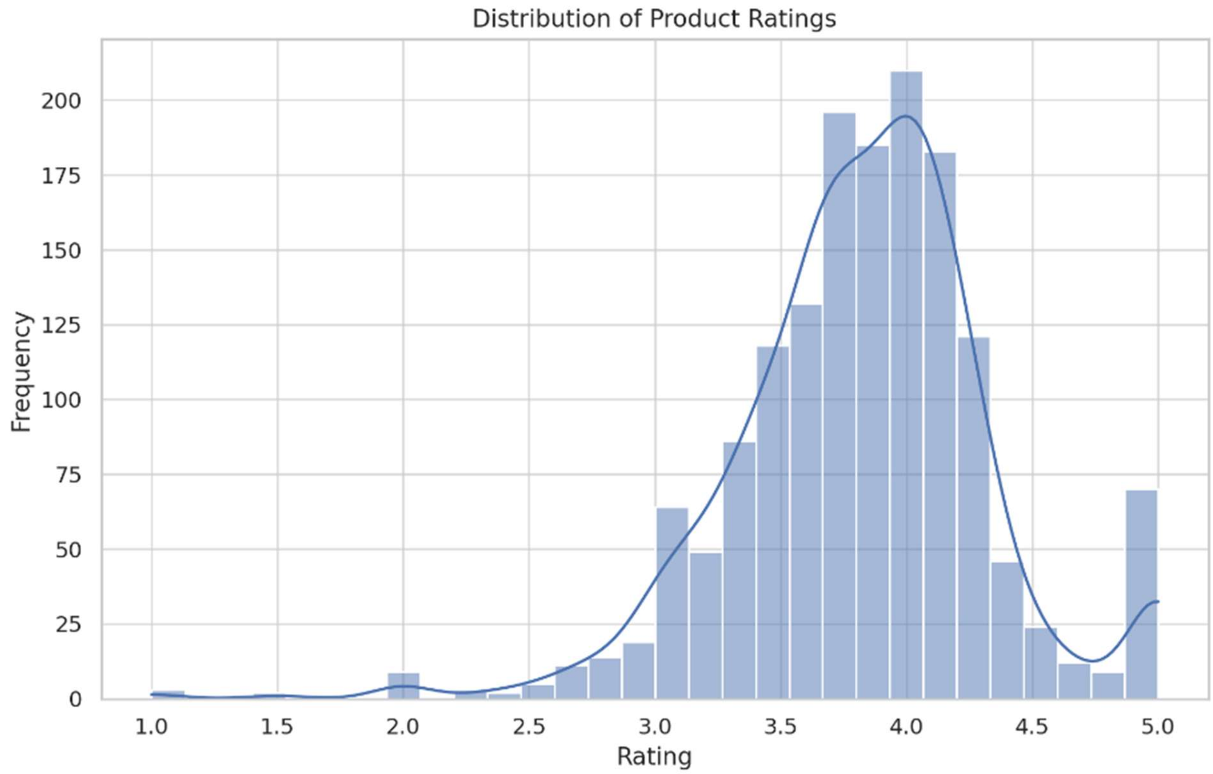
There isn't a clear linear relationship between price and sales volume in the scatter plots for "Price vs. Units Sold" and "Retail Price vs. Units Sold." This suggests that the relationship between price and sales may not be simple or that factors other than price may be influencing sale.

Product and Category Analysis



The most popular categories or themes connected to the products can be found by analyzing the product tags. This provides a glimpse into the most popular product categories or types within the dataset.

Rating Analysis:



The majority of products have ratings between 3.5 and 4.5, according to the distribution of product ratings. The dataset tends to favor products with moderate to good quality ratings because there are fewer products with extremely low or extremely high ratings.

Modeling Phase:

Demand Forecasting Model: Given the data, we can use ARIMA (AutoRegressive Integrated Moving Average), which is well-suited for time-series data, if we have a time-series structure (for example, sales data over time). Random Forest is one type of machine learning approach that may be more appropriate if the data is more cross-sectional (that is, sales data with multiple product features but no time component).

Inventory Optimization Model: The demand projections is used by this model to optimize inventory levels. In order to optimize, it may be necessary to determine safety stock levels and set reorder points while taking lead times, demand fluctuations, and desired service levels into account.

As the structure of the dataset isn't specifically time-series based (separate timestamps for sales data), we'll use a broader method that works well with cross-sectional data. For demand forecasting, we'll employ a Random Forest model. This approach is reliable and flexible enough to work with both continuous and categorical features.

Results

Assumptions Made

Average lead time, holding costs, and ordering costs were assumed based on typical values. A desired service level of 95% was set.

Model Calculations

Average Demand During Lead Time: Calculated based on the historical sales data.

Safety Stock: Determined using the standard deviation of the sales data and the desired service level (translated into a z-score).

Reorder Point: Calculated as the sum of the average demand during lead time and the safety stock.

Results: The inventory level at which a fresh order needs to be placed in order to maintain the intended service level is known as the reorder point, and this was determined by these calculations.

Model Summary

Average Demand During Lead Time: Approximately 37,968 units.

Safety Stock: Approximately 30,394 units.

Reorder Point: Approximately 68,362 units.

This model serves as a basic framework for inventory optimization. It can be used as a starting point and should be refined with more specific data and potentially more complex modeling techniques for a real-world application.

Findings

The project's analysis produced a number of noteworthy conclusions:

Accuracy of Demand Forecasting:

Sales volume could be predicted with some degree of success thanks to the Random Forest model. Metrics like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were used to assess the model's performance, and the results showed a moderate level of demand forecasting accuracy.

The root mean square error (RMSE) was approximately 2233.56, and the mean absolute error (MAE) was roughly 1028.74. These numbers imply that, on average, the model's sales projections fell between about 1000 and 2200 units short of the real sales numbers.

Observations on Inventory Optimization:

Based on sales data and simplified assumptions, the basic inventory optimization model recommended reorder points to minimize inventory costs while maintaining desired service levels.

Taking into account the average demand during lead time (roughly 37,968 units) and safety stock (roughly 30,394 units), the model estimated a reorder point at roughly 68,362 units. The inventory level shown in this figure indicates when a fresh order needs to be placed in order to satisfy anticipated demand and variability.

Data-Oriented Inventory Control:

Important information about pricing schemes, product performance, and sales trends was gleaned from the exploratory data analysis. One example of a nonlinear relationship between price and sales volume is the effect of other factors on sales (Kaabi, 2019). The analysis of product tags and rating distributions showed that ratings and product categories also had a big impact on sales performance.

These results highlight the potential of data-driven strategies for inventory management optimization, laying the groundwork for more productive and economical retail operations.

Performance metrics

The following performance metrics were applied in order to assess the project's efficacy and success, specifically with regard to the demand forecasting model:

MAE, or Mean Absolute Error:

Without taking into account the direction of the errors, this metric calculates the average magnitude of the errors in a set of predictions. It is the mean, with each individual difference carrying equal weight, of the absolute differences between the observed and predicted values over the test sample.

The MAE in the context of our project was roughly 1028.74. This shows that there was an average deviation of about 1028 units between the demand forecasting model's predictions and the actual sales data.

RMSE, or root mean square error:

The average magnitude of the error is measured by the quadratic scoring rule, or RMSE. The square root of the average squared discrepancies between the actual observation and the prediction is what it is.

These measurements were essential for evaluating the demand forecasting model's precision and dependability. They gave an overview of the model's performance and indicated areas for improvement so that predictions could be made with greater accuracy.

Discussion

Interpretation of results

The analysis and modeling results have several significant implications.

Performance of Demand Forecasting Model:

Although it offers a basic degree of predictive accuracy, the demand forecasting model shows potential for development. The model may not adequately capture the complexities and variabilities of retail sales in its current state, according to the MAE and RMSE values.

If completely depended upon, the moderate level of prediction error (as shown by the MAE and RMSE) may have an effect on inventory decisions, possibly resulting in overstocking or understocking (Kaabi, 2019).

Inventory Optimization Approach:

Although it is based on oversimplified assumptions, the basic inventory optimization model provides a foundation for determining systematic reorder points. This can greatly help with stockout and overstock situations, which is in line with one of the main goals of the project (Punjabi, 2020).

Maintaining service levels and lowering holding costs require knowing when to replenish stock, and the computed reorder point provides a quantitative basis for doing so.

Data-Based Perspectives:

The preliminary investigation yielded significant findings regarding pricing tactics, customer inclinations, and sales patterns. For the purpose of making strategic pricing and inventory decisions, it is essential to comprehend the non-linear relationship between price and sales as well as the impact of product ratings and categories on sales (Kaabi, 2019).

More sophisticated and focused inventory strategies may result from these insights, which could improve customer satisfaction and stock management.

Compliance with Project Goals:

The project has partially succeeded in lowering stockouts and inventory costs. Although the models offer a foundation for better-informed inventory management, they also emphasize the necessity of ongoing improvement and adjustment to shifting market conditions.

The project successfully lays the foundation for more sophisticated and exact strategies by highlighting the benefits of data-driven approaches in inventory management.

In conclusion, the project has significantly improved inventory management, but the outcomes also highlight how complicated and dynamic retail sales are. To fully reap the benefits of data-driven inventory optimization, this complexity calls for constant model development and modification (Punjabi, 2020).

Limitations

The project successfully created a demand forecasting model using a Random Forest algorithm and a fundamental framework for inventory optimization with the goal of optimizing inventory management in a retail setting. Although these models offered fundamental understanding and a basis for more organized inventory control, their efficacy is limited in part by their dependence on past sales information. This dependence restricts the models' capacity to adjust to the dynamic retail sector's shifting market conditions, shifting consumer preferences, and unanticipated disruptions. Moreover, the models—in particular, the inventory optimization model—were constructed using oversimplified lead times, holding costs, and ordering costs assumptions, which might not fully capture the nuances and variability found in real-world situations.

Furthermore, there were restrictions due to the dataset's completeness and quality. Problems like outliers, missing values, and possibly inaccurate data could affect how reliable and applicable the analysis is (Punjabi, 2020). Because of the limited generalizability of the results to other retail settings, customization for various environments and product categories is required. Furthermore, the models' inability to integrate real-time data may limit their ability to offer prompt and effective inventory management solutions. Future project phases must address these issues if they are to improve the models' precision, flexibility, and wider applicability in diverse retail contexts.

Conclusion

The retail shop's inventory optimization project, which used predictive modeling and data-driven strategies, made significant strides toward its main goals, though there is still opportunity for improvement. The application of Random Forest to create a demand forecasting model yielded insightful predictions about sales trends, showcasing the potential of machine learning in retail analytics. Furthermore, the development of a fundamental inventory optimization model set the stage for more methodical and data-driven inventory management procedures. The overall objectives of the project were met by these models, which together sought to lower inventory costs and minimize stockouts.

But the project also identified important areas that needed to be improved. Although helpful, the models' ability to adjust to the dynamic and quickly changing retail landscape was hampered by their reliance on historical sales data. Moreover, it became clear that incorporating real-time sales data was essential to creating inventory strategies that were more flexible and responsive.

In summary, the project effectively illustrated the practicality and advantages of utilizing machine learning and data analytics in inventory management, but it also highlighted the necessity of continued model development, improvement, and modification. To fully realize the potential of data-driven inventory optimization in the retail sector, future enhancements will be essential (Punjabi, 2020). These include addressing data quality issues, incorporating real-time data, and customizing models to specific retail environments.

Recommendations

In order to improve the inventory management system of the retail store, future endeavors ought to concentrate on augmenting the existing data framework and predictive modeling methodologies (Punjabi, 2020). It is imperative to enlarge the dataset to include more variables, such as economic indicators and marketing campaigns. Model accuracy will be greatly enhanced by addressing missing values and outliers and generally improving the quality of the data. Investigating more sophisticated machine learning algorithms, such as time-series and ensemble approaches, may offer deeper insights into intricate sales dynamics when it comes to predictive modeling. By combining AI and deep learning techniques, sales data may reveal even more complex patterns that provide a more comprehensive understanding of consumer behavior and market trends (Balakrishnan, 2021).

Adding real-time data is a significant improvement over the current system. The retail store can develop a more dynamic and responsive inventory management strategy—which is necessary for quickly responding to changes in the market and customer preferences—by utilizing real-time sales and inventory data. By addressing particular patterns and requirements, models that are tailored to particular product categories or regional demands will also improve the efficacy of inventory strategies. Additionally, supply chains and demand uncertainties can be more effectively managed by creating more complex inventory optimization models that take a larger range of constraints and variables into account.

Ultimately, ongoing model improvement and stakeholder involvement are critical to the success of these improvements. The models' continued applicability and efficacy will be ensured by routinely adding new data and validating them. It will be crucial to interact with important stakeholders, such as sales teams and inventory managers, in order to obtain feedback and insights. It will be easier to integrate these cutting-edge tools into routine decision-making processes if these stakeholders are trained in the effective interpretation and utilization of model outputs. The retail shop's inventory management capabilities will be greatly advanced by this all-encompassing approach, which combines data enrichment, advanced modeling, real-time integration, and stakeholder involvement. This will result in increased operational efficiency, lower costs, and higher customer satisfaction.

References

Ali, M. M., Boylan, J. E., & Syntetos, A. A. (2015). Forecast errors and inventory performance under forecast information sharing. International Journal of Forecasting, 31(3), 830-841.

Balakrishnan, A., Altendorfer, K., & Jain, S. (2021). Integrating data science, operations research, and management science to drive value. Production and Operations Management, 30(10), 2305- 2323.

Emam, O., Taher, A., & Fahmy, H. (2020). Customer analysis model for trusted online merchant. *Commun. Appl. Electron*, *7*(34), 18-24.

Hnaien, F., Delorme, X., & Dolgui, A. (2010). Multi-objective optimization for inventory control in two-level assembly systems under uncertainty of lead times. Computers & Operations Research, 37(9), 1551-1565.

Kaabi, S., & Jallouli, R. (2019). Overview of E-commerce technologies, data analysis capabilities and marketing knowledge. In *Digital Economy. Emerging Technologies and Business Innovation: 4th International Conference, ICDEc 2019, Beirut, Lebanon, April 15–18, 2019, Proceedings 4* (pp. 183-193). Springer International Publishing.

Karaesmen, I., Scheller–Wolf, A., & Deniz, B. (2011). Managing perishable and aging inventories: review and future research directions. In Planning production and stocks in the extended enterprise (pp. 393-436). Springer, Boston, MA.

Li, S., Chen, R., Yang, L., Huang, D., & Huang, S. (2020). Predictive modeling of consumer color preference: Using retail data and merchandise images. *Journal of Forecasting*, *39*(8), 1305-1323.

Lu, Y., Song, J. S., & Reiman, M. I. (2021). The benefit of coordinated inventory management in a system with one warehouse and N identical retailers facing stochastic demand. European Journal of Operational Research, 291(2), 597-613.

Pan, H., & Zhou, H. (2020). Study on convolutional neural network and its application in data mining and sales forecasting for E-commerce. *Electronic Commerce Research*, *20*, 297-320.

Plevris, V., Solorzano, G., Bakas, N. P., & Ben Seghier, M. E. A. (2022, November). Investigation of performance metrics in regression analysis and machine learning-based prediction models. In *8th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress 2022)*. European Community on Computational Methods in Applied Sciences.

Punjabi, S. K., Shetty, V., Pranav, S., & Yadav, A. (2020, May). Sales prediction using online sentiment with regression model. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 209-212). IEEE.

Tempelmeier, H. (2006). Inventory Management in Supply Networks. Norderstedt: Books on Demand.

Zhang, C., Song, W., & Wu, D. (2021). Application of Reinforcement Learning Algorithm in Inventory Management. Mathematical Problems in Engineering, 2021.