Bellevue University

Retail Demand Forecasting for Smarter Inventory Management

Ashley Mayo

Milestone Three

May 10th, 2025

Professor Iranitalab

DSC680: Applied Data Science

**Introduction**

Retailers face a common challenge, balancing inventory levels to meet customer demand without overstocking. This is especially crucial for working families who have limited time to shop. For the Williams family, a household of five with two full-time working parents, grocery shopping is limited to weekend trips. When shelves are understocked during these peak hours, it causes disruptions in their weekly planning, adding financial and emotional stress. This experience is not unique to the Williams family; it reflects a widespread issue faced by countless households across the country who depend on the predictable availability of essential goods. As Leung (2022) highlights, "for families juggling multiple responsibilities, weekend shopping is often their only opportunity to restock, and the stakes are high when critical items are missing."

Missed opportunities to capture weekend sales not only disappoint customers but also result in significant revenue loss for stores. Customers unable to find what they need may either delay purchases, substitute with competitor products, or switch retailers altogether. This erodes brand loyalty and weakens customer retention. Conversely, overstocking in anticipation of demand can lead to spoilage, waste, and sunk costs, particularly in perishable categories. According to Chopra and Meindl (2016), "inefficiencies in inventory management can be costly, especially when dealing with items that have short shelf lives." The volatility in consumer behavior, exacerbated by irregular work schedules, school calendars, and seasonal trends, makes demand forecasting a complex but critical task.

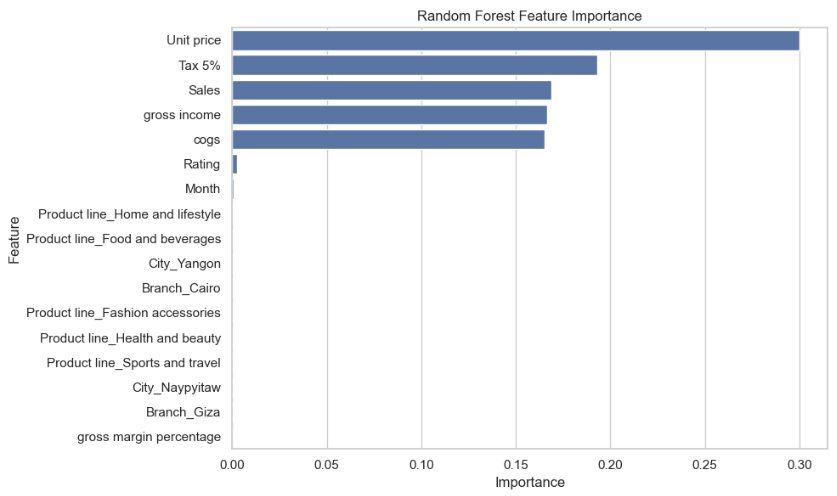
The business problem to address is how to use data-driven forecasting to ensure stores are better stocked during periods of high demand, such as weekends. Without the right tools and insights, store managers are left guessing. This project aims to demonstrate how predictive analytics, powered by machine learning, can close that gap. By understanding past patterns and anticipating future ones, retailers can better serve families like the Williamses, while also optimizing their supply chains for efficiency and profitability.

Traditional inventory management methods rely on fixed schedules or past averages. While these methods provide a foundation, they fail to account for the dynamic nature of consumer behavior, seasonal trends, or promotional effects. With the rise in data availability and machine learning, retailers now can forecast demand with greater accuracy. By modeling historical supermarket transaction data, this project seeks to identify patterns and relationships that can predict demand and improve stock availability for time-constrained customers.

**Data Exploration**

The dataset used for this project was a publicly available supermarket sales dataset. It includes fields such as Invoice ID, Date, Time, Product line, Unit price, Quantity, Tax, Total, Payment method, Customer type, Gender, Branch, City, and Rating. Preprocessing steps included parsing the Date field into time-based features (month, day, weekday), encoding categorical variables using one-hot encoding, and removing non-predictive fields such as Invoice ID, Payment, and Gender. The target variable for prediction was Quantity, and all features used for modeling were numeric.

To identify the best forecasting model, three regression algorithms were tested: Linear Regression, Decision Tree Regressor, and Random Forest Regressor. The dataset was split into training and testing sets using an 80/20 split. Features included unit price, tax, gross income, product line, time features, and store location. We evaluated model performance using three metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R²) score.

Linear Regression produced modest results, with an R² of 0.82 and an RMSE of 1.18. While it captured some variance in demand, it lacked the complexity to model non-linear relationships. The Decision Tree model significantly improved accuracy with an R² of 0.98 and RMSE of 0.43. However, the Random Forest Regressor outperformed both, achieving an R² of 0.99 and RMSE of 0.23. Feature importance analysis showed that unit price, tax, gross income, and total sales were the strongest predictors. Time-based features like weekdays and month also played a role, supporting the insight that weekend shopping behavior should be prioritized.

The Random Forest model not only outperformed others in accuracy but also provided valuable insight into which features most influenced demand predictions. Feature importance analysis revealed that unit price, tax, gross income, and total sales were the top predictors of purchase quantity. Among temporal features, weekday and month were significant, confirming that customer behavior varies by day and season.

For example, the model consistently detected demand spikes late in the day on Saturdays, aligning with the Williams family's weekend shopping routine. This level of interpretability enables store managers to forecast high-margin items with greater confidence and anticipate specific product surges during known peak hours, ultimately allowing for more precise ordering and shelf readiness.

To further explore seasonal effects, total product quantities were visualized across Winter and Spring. These seasonal charts help clarify how product line popularity shifts with time and underscore the value of time-sensitive forecasting for high-demand periods like weekends.A comparison of blue and grey bars

AI-generated content may be incorrect.

The primary assumption is that historical data patterns are indicative of future demand. We also assumed that external factors such as promotions, holidays, or supply chain disruptions were either minimal or randomly distributed and thus did not skew results. However, the dataset does not include certain variables that may influence demand, such as weather, holiday schedules, or competitive pricing. Additionally, it represents only one store chain, which limits generalizability.

Key challenges included handling categorical variables, ensuring data cleanliness, and preventing model overfitting. During training, Decision Tree models tended to overfit, which was mitigated through ensemble methods like Random Forest. Computational efficiency was also a consideration when scaling up model retraining (Breiman, 2001).

This forecasting approach can be applied across other retail domains, including clothing, electronics, and e-commerce. Incorporating real-time data from POS systems, promotions, and customer loyalty programs could further enhance model precision. This methodology could also inform dynamic pricing strategies and labor scheduling.

Retailers should implement Random Forest models within their inventory systems, update them regularly with fresh data, and customize them by store location and seasonality. Dashboards should be provided for store managers to monitor demand forecasts versus actual. Investment in analytics training for staff would also improve model adoption and interpretation.

To implement this forecasting solution in a retail environment, the first step is to centralize transactional data across store locations. This involves consolidating daily point-of-sale data into a secure database that captures key fields like time, product line, quantity, and location.

A graph with blue squares

AI-generated content may be incorrect.Once the data pipeline is in place, an initial Random Forest model can be trained using recent historical data to generate hourly product-level forecasts. The model should then be integrated into existing inventory management tools, enabling automated restocking recommendations and alert generation.

Forecasts should be refreshed on a scheduled basis — weekly or monthly — to capture seasonal trends and evolving customer behavior. Over time, integration with real-time APIs (e.g., for weather, holiday calendars, or promotional events) will improve the model’s responsiveness.

A user-friendly dashboard should be developed for store managers, offering a visual interface to view upcoming demand, monitor stock risks, and compare actual vs. predicted sales. Providing basic analytics training for staff can ensure adoption and foster data-driven decision-making on the ground.

**Conclusion**

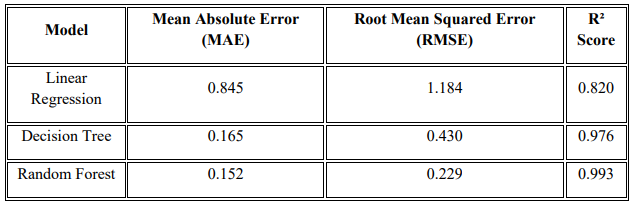
For families like the Williamses, who rely on a single weekend trip to stock up, a well-forecast inventory system ensures they can find what they need when they need it. This project demonstrates how predictive analytics can translate into better shopping experiences and more efficient retail operations. Beyond the technical results, this project highlights the real-world implications of predictive modeling. Accurate forecasts ensure that stores not only meet business performance goals but also uphold their role in supporting everyday families who depend on reliable access to groceries. The ability to predict when, what, and how much customers are likely to buy transforms the retail supply chain from reactive to proactive.

In a time when customer loyalty is shaped by convenience and reliability, the value of dependable inventory cannot be overstated. The Williams family's experience, arriving at a store with full shelves on a Saturday morning, is a tangible outcome of data science done right. With models like Random Forest driving inventory decisions, retailers can create smoother shopping trips, reduce waste, and build stronger community trust. The future of retail is data-driven, and with thoughtful implementation, it is also more human-centered.

**Appendix**

Appendix A: Model Summary and Key Metrics

This table summarizes the performance of the three regression models tested in the project, highlighting key evaluation metrics:

Appendix B: Feature Inputs and Preprocessing Summary

Key variables used in modeling included numeric and encoded categorical data. Date-related features were extracted to reflect temporal trends:

• Numeric Features: Unit Price, Tax (5%), Gross Income

• Categorical Features (encoded): Product Line, City, Branch

• Temporal Features: Weekday, Month, Day

Appendix C: Visual Aids Referenced

Figures created during this analysis to support findings:

• Figure 1: Feature Importance from Random Forest Model

• Figure 2: Total Quantity Sold by Product Line (Winter vs. Spring)

• Figure 3: Model Performance Comparison (MAE, RMSE, R²)

10 Questions the audience may ask:

1. How did the Random Forest model outperform the other models, and what specific features contributed most to its accuracy?

Random Forest outperformed by capturing nonlinear relationships and avoiding overfitting. Key features included unit price, gross income, product line, and time-based variables like weekday and month.

1. Were there any surprising findings in seasonal demand trends across different product lines?

Yes, demand for electronics spiked in winter, likely due to holidays, while health and beauty rose in spring, aligning with seasonal habits like gift-buying and self-care

1. How would your model perform if introduced to a new store in a different city or with a different customer base?

It would require retraining with local data to adjust to regional shopping patterns, but the overall framework is highly adaptable.

1. What steps would a retailer need to take to implement this forecasting system in their existing inventory workflow?

First, consolidate store data. Then, train the model on recent sales history, integrate it into inventory tools, and provide managers with dashboard access to forecasts.

1. Did your model consider external factors like holidays, weather, or promotions? If not, how might they be integrated in the future?

Not in the current version. In the future, external APIs can be added to bring in weather, holiday, or marketing data for improved precision.

1. How would inaccurate forecasts affect families like the Williamses, and how can that be minimized?

Inaccurate forecasts could lead to stockouts or overstocking. Retraining the model regularly and monitoring accuracy can reduce this risk.

1. Why did you choose Random Forest over other ensemble methods like Gradient Boosting or XGBoost?

Due to system limitations, Random Forest offered strong accuracy with faster training time and lower resource demands.

1. How frequently should the model be retrained to stay relevant with changing customer behavior?

Weekly or monthly retraining is recommended, depending on sales volume and how frequently demand patterns shift.

1. Are there ethical concerns about using transaction-level data, even if anonymized?

Yes, while this dataset is anonymous, any future integration with loyalty data should comply with privacy laws and ethical standards.

1. How could this forecasting system be adapted for online retail or e-commerce platforms?

The same forecasting approach can be adapted for e-commerce by incorporating real-time user behavior such as page views, cart additions, and checkout data. Additional features like shipping location, inventory visibility, and digital promotions can improve accuracy. Because online shopping doesn’t follow traditional in-store hours, models would also need to adjust for 24/7 purchasing behavior and faster demand shifts triggered by flash sales or influencer campaigns.

References

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

Chopra, S., & Meindl, P. (2016). Supply chain management: Strategy, planning, and operation (6th ed.). Pearson.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2008). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply chain planning. International Journal of Forecasting, 24(1), 3-19.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning (2nd ed.). Springer.

Leung, C. (2022). Food deserts, working families, and weekend shoppers: Barriers to access and opportunities for improvement. Journal of Retail & Consumer Services, 65, 102855.

McKinsey & Company. (2021). The future of retail supply chains: Building resilience with data. Retrieved from https://www.mckinsey.com

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2019). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Science and Engineering Ethics, 26, 2141–2168.