Sortive: Intelligent Product Grouping and Layout Recommendation Based on Transaction Data Using FP-Growth Algorithm

Justine Delos Reyes Gravamen[1], Jerameel Osorio Udtujan[2], Jenny Rose Irish Resma Marasigan[3], Prof. Anna Liza O. Villanueva[4], Prof. Marvin H. Bicua[5]

Abstract
Small to mid-scale supermarkets in the Philippines typically rely on intuitive product placement and supplier-based logistics, often overlooking latent purchasing patterns within their transaction data. This study developed Sortive, an intelligent system utilizing the FP-Growth algorithm to transform raw Point-of-Sale (POS) records from the JMFAITHHOPE Store into actionable spatial strategies. Following the Iterative Waterfall Model, the system architecture features a Decoupled Data Pipeline composed of four functional modules: User Access and Security, Data Ingestion and ETL, an FP-Growth Analytical Engine, and a Layout Optimization Dashboard. Technical evaluation by IT experts based on the ISO/IEC 25010 product quality model yielded a Grand Weighted Mean of 3.41 (Effective), with a peak score in Security (3.62) validating the robustness of the system's role-based access controls. User acceptance testing by store personnel resulted in a mean of 3.30 (Strongly Agree) for Functionality, specifically highlighting the value of narrated, data-driven placement insights (3.50). The findings prove that algorithmic Market Basket Analysis successfully replaces the traditional "Manual Loop" management practice with statistically-proven layout strategies, offering an affordable and scalable decision-support tool for local retail optimization.

INTRODUCTION

In the Philippine retail industry, many small to mid-scale supermarkets continue to rely on traditional shelf layouts and manual product placement driven by managerial intuition and supplier-based logistics rather than data. As illustrated by the "Manual Loop" observed at the JMFAITHHOPE Store, these static arrangements often overlook purchasing trends and evolving customer behaviors, resulting in inefficient store navigation and missed revenue opportunities. While Point-of-Sale (POS) systems capture rich transaction details, this data frequently remains "un-mined" and stagnant, serving only for basic auditing rather than strategic optimization. Kholod, Celani, and Ciaramella (2024) emphasize that transaction-level analytics are essential for enhancing demand forecasting and customer engagement, yet local retailers often lack the technical infrastructure to transform raw logs into actionable intelligence.

Market Basket Analysis (MBA) provides a solution by uncovering hidden relationships between products frequently purchased together. However, traditional methods like the Apriori algorithm can be computationally expensive due to multiple database scans, especially in retail environments with high SKU diversity. In contrast, the Frequent Pattern Growth (FP-Growth) algorithm compresses data into a tree structure, allowing for faster and more efficient mining. Idris et al. (2022) found that FP-Growth significantly outperforms Apriori in computation time, making it a preferred choice for real-time recommendation engines. Recent implementations in Python-based systems by El-Houssainy et al. (2024) and Nguyen et al. (2023) further validate that FP-Growth can provide high-speed decision support within accessible web platforms.

Despite these advancements, a significant gap persists: most analytical solutions are designed for large-scale hypermarkets, leaving small Philippine retailers underserved due to high costs and technical complexity. As observed by Isharyani et al. (2024), small retailers often rely on improvised technology setups, requiring tools that prioritize low complexity and compatibility with existing manual workflows.

To address this gap, the present study developed **Sortive**, an intelligent system designed to empower small to mid-scale supermarkets to harness their existing POS data for spatial optimization. Sortive utilizes the FP-Growth algorithm to translate technical metrics—Support, Confidence, and Lift—into human-readable, data-driven placement insights that assist managers in evidence-based decision-making.

METHODS

This study utilized a Descriptive-Developmental research design with a quantitative approach to develop and evaluate the Sortive system. The descriptive aspect analyzed the existing "Manual Loop" management practices and stagnant data utilization processes of the JMFAITHHOPE Store, while the developmental aspect focused on creating an intelligent recommendation system guided by the Iterative Waterfall Model of the Software Development Life Cycle (SDLC).

The study involved two purposely selected participant groups for the evaluation phase: a group of IT experts who assessed technical effectiveness, and supermarket managers and staff from local small-to-mid-scale supermarkets who evaluated user acceptance. Purposive sampling was used to select supermarkets that utilize a digital POS system with 1-3 active checkout counters. Data were gathered through researcher-made Likert-scale surveys aligned with the ISO/IEC 25010 Product Quality Model, assessing technical attributes such as functionality, reliability, efficiency, compatibility, and security, alongside practical user experience metrics.

The technical procedure centered on the development of a Decoupled Data Pipeline. Raw Point-of-Sale (POS) transaction records were processed through an ETL (Extract, Transform, Load) module built in Python, which cleaned and mapped product SKUs into general categories (e.g., Dairy, Bakery) to form binarized "baskets" suitable for analysis. These baskets were processed by the FP-Growth algorithm using the mlxtend and pandas libraries to identify frequent itemsets and generate association rules based on support, confidence, and lift metrics. The system architecture employed a three-tier structure: a web-based frontend (HTML/CSS/JS), a Flask server backend, and a lightweight, file-based SQLite database for portable data management. This methodology ensured that the system remained computationally efficient for resource-constrained environments while providing actionable, data-driven layout recommendations.

RESULTS AND DISCUSSION

This section presents the gathered data analyzed and interpreted to determine the technical effectiveness and user acceptance of Sortive. The framework of the analysis and interpretation is guided by the specific research questions established in the introduction, focusing on the system's capacity to transform raw retail records from the JMFAITHHOPE Store into actionable spatial optimization strategies using the FP-Growth algorithm

Table 1. Summary of System Effectiveness (IT Experts)

Criteria	Average Weighted Mean	Verbal Interpretation
Functionality	3.37	Strongly Agree
Reliability	3.33	Strongly Agree
Efficiency	3.41	Strongly Agree
Compatibility	3.37	Strongly Agree
Security	3.62	Strongly Agree
Usability	3.37	Strongly Agree
Grand Weighted Mean	**3.41**	**Effective**

Table 1 presents the IT experts' overall evaluation of Sortive's effectiveness based on ISO/IEC 25010:2023 criteria, yielding a grand weighted mean of 3.41 (Effective). The highest-rated characteristic was Security (3.62, Strongly Agree), confirming that the Role-Based Access Control (RBAC) and non-repudiable audit trails provide robust protection for sensitive retail data. This is followed by Efficiency (3.41, Strongly Agree), which validates the Flask and SQLite architecture's ability to process large transaction datasets within reasonable timeframes. Lower but still positive ratings for Reliability (3.33) and Functionality (3.37) confirm that the system correctly executes the FP-Growth pipeline and handles operational errors gracefully without critical failures. These results demonstrate that Sortive meets high technical quality standards and is a reliable tool for automated pattern discovery in supermarket environments.

Table 2. Summary of User Acceptance (Supermarket Personnel)

Criteria	Average Weighted Mean	Verbal Interpretation
Usability	3.18	Agree
Functionality	3.30	Strongly Agree
Reliability	3.25	Agree
Efficiency	3.37	Strongly Agree
Grand Weighted Mean	**3.28**	**Highly Accepted**

Table 2 summarizes the supermarket personnel's evaluation of Sortive's user acceptance, yielding a grand weighted mean of 3.28 (Highly Accepted). The highest-rated characteristic was Efficiency (3.37, Strongly Agree), indicating that the end-users found the system responsive, with pages loading quickly and analysis results appearing shortly after data upload. The Functionality score of 3.30 (Strongly Agree) confirms the practical utility of the "Actionable Insights" module, which received a specific indicator score of 3.50 for providing valuable spatial recommendations. While Usability (3.18) and Reliability (3.25) received positive ratings, the findings suggest that while technical metrics like support and confidence require narration, the overall dashboard layout effectively assists store managers in making data-driven placement decisions for the JMFAITHHOPE Store.

CONCLUSIONS AND RECOMMENDATIONS

The study concludes that Sortive successfully bridges the critical gap between raw, stagnant transaction logs and actionable spatial strategies for small to mid-scale supermarkets. Multi-stakeholder evaluation based on ISO/IEC 25010:2023 standards confirmed this effectiveness, with IT experts rating technical quality as Effective (grand mean 3.41) and store personnel demonstrating High Acceptance (grand mean 3.28) . The implementation of a "Decoupled Data Pipeline" was found to be a vital architectural requirement for high-availability retail environments, as it ensures data integrity by strictly isolating the ETL process from the core analytical engine. Furthermore, the system effectively replaces the traditional, intuition-based "Manual Loop" of the JMFAITHHOPE Store with a statistically-proven "Knowledge Discovery" pipeline utilizing the FP-Growth algorithm for efficient frequent pattern mining. Ultimately, the research proves that providing narrated, data-driven placement insights significantly improves upon intuitive layouts by accurately matching physical shelf organization with statistically-proven consumer purchase patterns.

To ensure the long-term technical evolution of the system, it is recommended that future researchers and system developers integrate an API layer for Direct Point-of-Sale (POS) Connection. This would eliminate the need for manual CSV uploads and provide store managers with real-time analytical updates on purchasing trends by developing an automated data-fetching service that pulls transaction logs directly from the store's backend database into the Sortive ETL staging tier. Additionally, supermarket owners and UI/UX designers should focus on the re-incorporation and expansion of visual illustrations for movable racks to provide floor staff with a dynamic physical guide for shelving execution beyond fixed category adjustments. This can be achieved by integrating a 2D layout simulator into the main dashboard that visualizes recommended rack placements based on seasonal shifts in product associations. Finally, local retail analysts and store owners are encouraged to implement Predictive Demand Modeling to refine "Optimization Potential" metrics by accounting for external demand factors such as payday cycles and promotional events. This enhancement involves pairing FP-Growth association results with time-series forecasting libraries to anticipate fluctuations in co-purchase patterns before they occur.

REFERENCES

1. Aithal, R. K., Choudhary, V., Maurya, H., Pradhan, D., & Sarkar, D. N. (2023). Factors influencing technology adoption amongst small retailers: Insights from thematic analysis. *International Journal of Retail & Distribution Management*, *51*(1), 81–102. https://doi.org/10.1108/IJRDM-05-2021-0277
2. Andrada, J., Amores, E., Cabaluna, J., Carpio, E., & Zuasula, R. (2020). Facility layout design for independent grocery store using systematic layout planning and house of quality. *International Journal of Industrial Engineering and Operations Management*, *2*(4), 56–63.
3. Czerniachowska, K., & Subbotin, A. (2021). Merchandising rules for effective shelf space allocation in retail stores. *Journal of Retail Marketing and Management*, *17*(2), 101–113. https://doi.org/10.15611/jrmm.2021.17.2.101
4. Farheen, Z., & Dharani, A. (2024, November). Prediction of customer purchasing patterns for retail optimization using market basket techniques. In *2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)* (pp. 1–5). IEEE.
5. Idris, A. I., Sampetoding, E. A., Ardhana, V. Y. P., Maritsa, I., Sakri, A., Ruslan, H., & Manapa, E. S. (2022). Comparison of Apriori, Apriori-TID and FP-Growth algorithms in market basket analysis at

grocery stores. *The IJICS (International Journal of Informatics and Computer Science)*, *6*(2), 107–112.
6. Isharyani, M. E., Sopha, B. M., Wibisono, M. A., & Tjahjono, B. (2024). Retail technology adaptation in traditional retailers: A technology-to-performance chain perspective. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(1), 100204. https://doi.org/10.1016/j.joitmc.2024.100204
7. Moreno Arboleda, J., Díaz, L., & Restrepo, A. (2024). Inventory optimization through market basket analysis: A data-driven approach to reduce stockouts. *Journal of Retail Analytics*, *10*(3), 45–56.
8. Pietrini, R., Paolanti, M., Mancini, A., Frontoni, E., & Zingaretti, P. (2024). Shelf Management: A deep learning-based system for shelf visual monitoring. *Expert Systems with Applications*, *255*, 124635. https://doi.org/10.1016/j.eswa.2023.124635
9. Rak, T., & Żyła, R. (2022). Using data mining techniques for detecting dependencies in the outcoming data of a web-based system. *Applied Sciences*, *12*(12), 6115. https://doi.org/10.3390/app12126115
10. Shawkat, M., Badawi, M., El-Ghamrawy, S., Arnous, R., & El-Desoky, A. (2022). An optimized FP-growth algorithm for discovery of association rules. *The Journal of Supercomputing*, *78*(4), 5479–5506. https://doi.org/10.1007/s11227-021-03908-7
11. Siljamäki, A. (2022). A decision support approach to achieve competitive advantage for a hypermarket chain. *Journal of Business Economics*, *92*(5), 809–827. https://doi.org/10.1007/s11573-021-01084-2
12. Sri Darshan, M., Nithinraj, N., & Jaisachin, B. (2024). Integrating data mining and predictive modeling techniques for enhanced retail optimization. *International Journal of Computer Science and Information Security (IJCSIS)*, *22*(5).
13. Tiainen, M. (2021). *Forecasting seasonal demand at the product level in grocery retail*. [Unpublished manuscript/Thesis, University Name].
14. Vogel, C., Crozier, S., Penn-Newman, D., Ball, K., Moon, G., Lord, J., ..., & Baird, J. (2021). Altering product placement to create a healthier layout in supermarkets: Outcomes on store sales, customer purchasing, and diet in a prospective matched controlled cluster study. *PLoS Medicine*, *18*(9), e1003729. https://doi.org/10.1371/journal.pmed.1003729
15. Xia, L., & Ma, H. (2025). Enhancing retail store layout and recommendations through market basket analysis. *International Journal of Data Mining & Business Intelligence*, *15*(1), 21–33.
16. Yoon, S., Lee, H., & Koo, D. (2020). Effects of store layout on perceived crowding and shopping intentions: The moderating role of gender. *Journal of Retailing and Consumer Services*, *54*, 102003. https://doi.org/10.1016/j.jretconser.2019.102003
17. Zhao, J., & Wang, S. (2025). Data-driven shelf stocking: Robustness and vertical-location effect. *Production and Operations Management*, *34*(2), 242–261. https://doi.org/10.1111/poms.14079