Predictive Ordering: Using Time-Segmented Association Rule Mining to Reduce Food Surplus and Improve Savings in Cafés

Caroline Bennett and Ryan Zambrano
College of Computer Science, Florida State University
CIS 4930: Data Mining
Dr. Yushun Dong
December 4, 2025

Predictive Ordering: Using Time-Segmented Association Rule Mining to Reduce Food Surplus and Improve Savings in Cafés

According to a food waste report published by ReFED in 2025, 85% of the United States' surplus food in 2023 goes to waste destinations. That percentage equates to 62.815 million tons in pounds. Surplus food is defined as all food that goes unsold or unused by a business, home, or restaurant, including food that is donated or repurposed. Food waste, on the other hand, is the subset of surplus food that is sent to disposal destinations like landfills, composting or the sewer (ReFED, 2025).

Within the retail food sector, small businesses, such as independent cafés, often struggle with food surplus due to difficulties in accurately predicting inventory, leading to excessive purchasing of food items. Consequently, these businesses suffer from significant financial losses (ReFED, 2025). The coffee shop sector is a competitive industry where sales are a crucial business activity that can significantly boost profitability (Wahyuningsih & Prasetyaningrum, 2023). In such a rapidly growing and competitive landscape, businesses must maintain product quality, anticipate customer demands and identify top-selling products to ensure efficiency and effectiveness. Unfortunately, operational challenges exist: customers frequently inquire about an extensive selection of menu options, which can cause transactions to take longer than desired, potentially leading to customer dissatisfaction and reduced turnover. This inefficiency is compounded by the difficulty in accurately predicting inventory, resulting in excessive surplus and significant financial losses. While this issue permeates the food industry within the United States, this research project narrows the scope down to cafés. The goal of this research project is to provide data-driven results that can help the retail food industry reduce lost revenue and decrease the amount of food surplus produced by the retail food sector.

In efforts to work towards a solution, this project focuses on optimizing inventory by analyzing customer purchasing behavior. Data mining is an analytical technique used to extract new information and valuable knowledge from large datasets by using statistics, artificial intelligence and machine learning. This approach is particularly effective for discovering hidden patterns, such as association rules, which identify items frequently bought together. The strength and relevance of these rules are evaluated by three key measurements: support, confidence and lift. By identifying strong association rules, items frequently bought together, and segmenting these patterns by the time of day, cafés can transition from generalized purchasing to targeted, predictive ordering. For example, if the rule discovered, {Iced Latte} → {Chocolate Croissant}, has high confidence only between 6:00 am and 8:00 am, the café can reduce its almond croissant inventory purchased for later in the day, leading to savings and a reduction in food surplus. This research study aims to operationalize association rule mining to deliver clear insights that directly contribute to financial health and, ultimately, improved sustainability. The business world is experiencing rapid growth and fierce competition, especially within the coffee shop industry. As the business landscape evolves, it becomes essential for entrepreneurs to explore various strategies to maintain product quality, anticipate customer demands and identify top-selling products.

Literature Review

The foundation of this project is built on Association Rule Mining, a key data mining application introduced in 1993. As noted by Abubakar (2015), association rules "are used to identify relationships among a set of items in a database" and aim "to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories” (p. 173). This concept was initially motivated by market basket analysis, which allows companies to "more fully understand purchasing behavior and, as a result, better target market audiences" (p. 173). Market-basket analysis utilizes sales promotion, shelf management and inventory management to produce those results that directly target the market audiences (Tan et al., 2024).

The main algorithm used for this task is the Apriori Algorithm: a "major algorithm used in mining frequent itemsets in a large data set" (Abubakar, 2015, p.173). The Apriori algorithm focuses on association analysis and determines the association between frequently occurring data. Association Rule Mining follows a two-step approach, beginning with frequent itemset generation followed by rule generation. Three types of measures determine the strength of these rules: 1) Support, which evaluates how often an itemset, two items bought together, appears in the dataset; 2) Confidence. which evaluates the probability that one item, ‘Y,’ is present in a transaction, given that another item, ‘X,’ was previously ordered; 3) Lift, which evaluates how much more likely item ‘X’ is to be bought when item ‘Y’ is bought randomly (Abubakar, 2015). The goal is to identify all rules that meet user-specified support, confidence and lift thresholds.

While data scientists have well-established the capability of Association Rule Mining for transactional data, current food surplus research tends to focus on broad technological solutions rather than specific data mining techniques. For example, research shows that AI-based systems have successfully helped hospitality establishments decrease food waste by 23% to 51% (Sigala et al., 2025). Furthermore, research describes the antecedent causes of food waste, such as general inventory and overproduction issues (Mercier, 2025). Despite this evidence, a notable research gap remains: there is little to no work applying data analytics or time-segmented association rule mining to decrease food surplus in small-scale, high-turnover settings like cafés. Our project aims to bridge this gap by applying Association Rule Mining to provide time-relevant insights into inventory.

Methology

The research project utilized Apriori Association Rule Mining, a foundational technique in Market Basket Analysis, to discover statistically significant and commercially actionable item co-occurrence patterns within the café’s transaction data.

To account for the primary assumption that customer purchasing habits are highly dependent on the time of day, a critical step involved time segmentation of the transaction data before rule generation. Transactions were categorized into four distinct time segments (Morning Weekday, Morning Weekend, Afternoon Weekday and Afternoon Weekend) to enable time-specific rule discovery and optimal inventory suggestions.

The Apriori Algorithm was employed to discover frequent itemsets and generate strong association rules efficiently. This process utilizes an iterative search where *k*-itemsets are used to explore (*k*+1)-itemsets. The core efficiency of the algorithm relies on the Apriori property, which states that all non-empty subsets of a frequent itemset must also be frequent. This principle is used to prune the search space, significantly reducing computation. Once frequent itemsets were established, association rules in the format {A} → {B} were generated and evaluated based on support, confidence and lift (Abubakar, 2015). The implementation was conducted using Python for efficient execution of the iterative algorithm.

The Apriori algorithm was used over more computationally efficient modern alternatives, such as FP-Growth or Eclat, due to its simplicity, transparency and the direct interpretation of its rule-generation process, which is essential for business stakeholders who must understand the logic behind operational changes (Tan et al., 2019). A significant, well-documented fault of the Apriori algorithm is the "candidate generation bottleneck," where the algorithm requires multiple scans of the transactional database to count support for itemsets, potentially resulting in high computational costs on massive datasets. This limitation, however, is effectively addressed in the context of this study by the manageable scale of a single café's transaction history and the rigorous application of the Apriori property, which prunes the search space by discarding infrequent itemsets early. Ultimately, Apriori is preferred over opaque "black box" models because its transparent "if-then" logic creates trust and actionability for management, ensuring that inventory decisions are based on clear, verifiable customer behaviors rather than abstract mathematical probabilities.

Implementations

To ensure that the generated rules were both statistically valid and commercially relevant, strict minimum thresholds were applied across all time segments:

I. Minimum Support: A threshold of 2% or greater was set to ensure that the pattern is frequent enough within the overall dataset to warrant operational changes.

II. Minimum Confidence: A threshold of 40% or greater was set to ensure that the predictive reliability of the rule is high, providing management with high-confidence predictions.

III. Rule Structure Assumption: The analysis focused exclusively on simple one-to-one antecedent-consequent rules, {Item A} → {Item B}, which maximizes the actionability of the findings for staff training and physical placement.

Experimental Results and Discussion

The time-segmented analysis successfully isolated a single, highly dominant association rule, {Ouro Brasileiro shot} → {Ginger Scone}, whose key metrics varied significantly across time segments. No strong patterns (Confidence $\geq$ 40%) were observed during the designated "Evening" segment (post-4 pm).

Results

Core Rule Metrics

The metrics confirm the pattern's strength, reliability and frequency across the strong segments. As shown in the Association Rule Mining Dashboard (Figure 1), the pattern shows universally high performance across all evaluated metrics.

Figure 1

Association Rule Mining Dashboard Ouro Brasileiro shot → Ginger Scone.

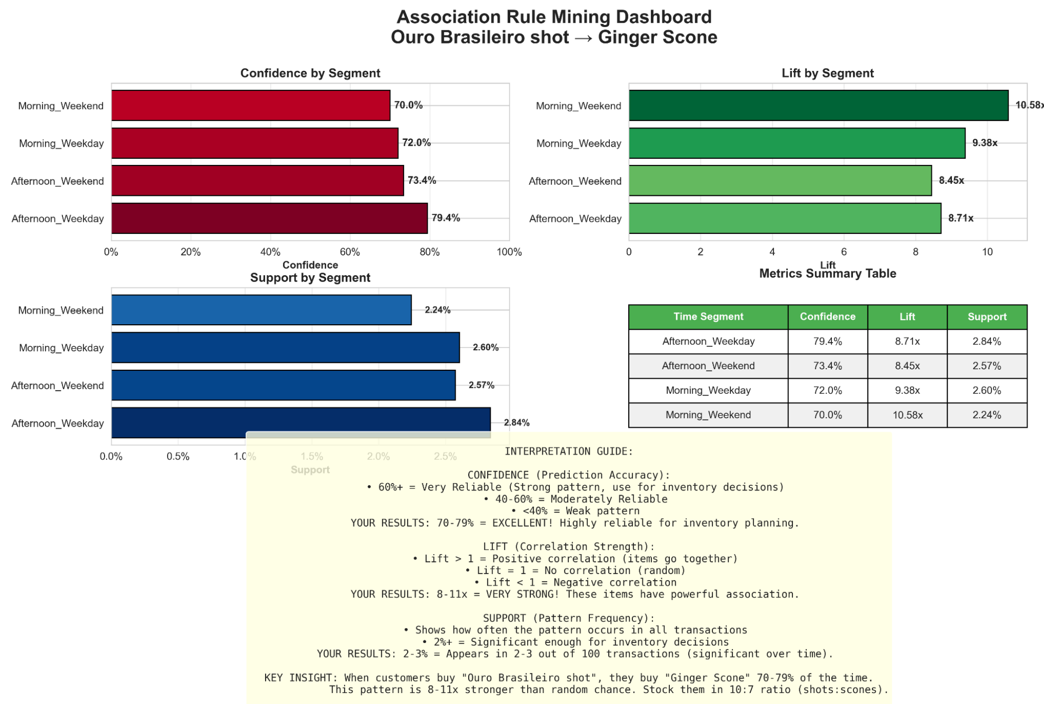

Table 1

Figure 1 Interpretation

Metric	Range of Results	Interpretation
Support	2.2% - 2.8%	The pattern occurs frequently enough (2-3 out of 100 transactions) to be highly significant for inventory decisions.
Confidence	70.0% - 79.4%	This represents high predictive accuracy: approximately 8 out of 10 customers who purchase the antecedent item also purchase the consequent item.
Lift	8.71x- 10.58x	The purchase of the Ginger Scone is 8 to 10 times more likely than random chance when the shot is purchased, confirming a strong positive correlation.

Time-Segmented Analysis and Inventory Strategy

The results directly inform a time-based inventory strategy, referred to as the 10:7 Rule, which dictates the optimal stocking ratio of Ginger Scones relative to Ouro Brasileiro shots. The Recommended Inventory Stocking Ratios chart (Figure 2) and Table 2 describe its implementation.

Figure 2

Recommended Inventory Stocking Ratios Based on Association Rule Confidence.

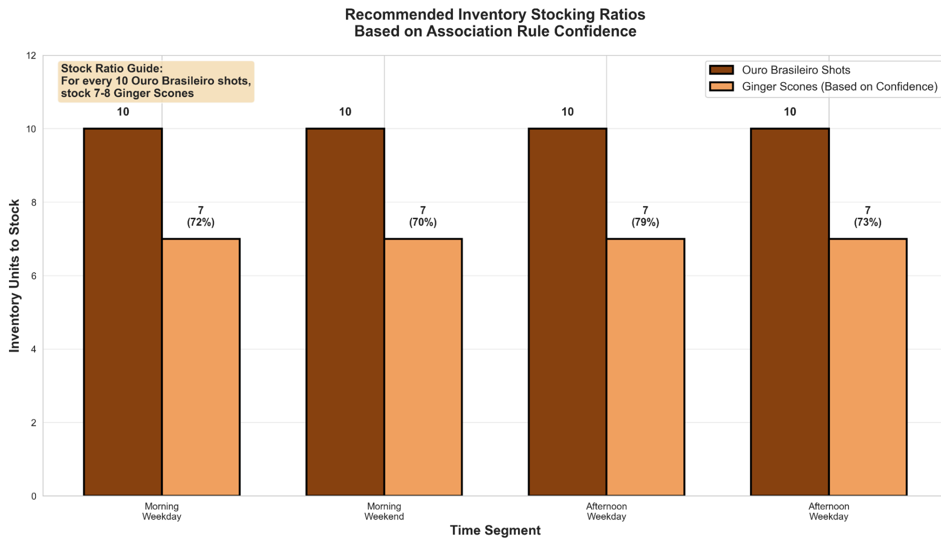

Table 2

Time-Segmented Performance and Dynamic Inventory Strategy

Time Segment	Confidence	Lift	Actionable Ratio (Shot:Scone)	Impact on Inventory Planning
Afternoon Weekday	79.4%	8.71x	10:8	Highest confidence and most reliable segment for prediction: Stock 7.9 scones per 10 shots.
Morning Weekday	72.0%	9.38x	10:7	High ratio required: Stock 7.2 scones per 10 shots.
Morning Weekend	70.0%	10.58x	10:7	Highest lift and strongest correlation: Stock 7.0 scones per 10 shots.
Evening (both)	< 40%	N/A	Conservative Stocking	The pattern disappears completely: Stock significantly fewer scones (50% of shot inventory) to reduce surplus.

Discussion

The results demonstrate the efficacy of time-segmented mining. The Afternoon Weekday segment yielded the highest confidence (79.4%), indicating that this is the most reliable time for staff to execute an upselling strategy. On the other hand, the Morning Weekend segment showed the highest Lift (10.6x), revealing the strongest non-random correlation, though with a slightly lower absolute confidence. Additionally, the absence of any strong rule in the Evening segment provides a basis for an adjustment to reduce surplus.

The data suggests that stocking Ginger Scones based on the high ratio is only justified during morning and afternoon hours. This finding transitions inventory management from a static daily estimate to a dynamic, time-dependent model, addressing the problem of food surplus during non-peak hours.

Operational Recommendations

The high confidence and time dependency of the rule provide three key operational recommendations, directly derived from the data's findings:

I. Staff Training and Upselling: Train staff on the high reliability of the pattern. This includes suggestions such as 'Would you like a Ginger Scone with that?' which is predicted to have a success rate between 70% and 79.4%. This leverages the high confidence metric to increase the average transaction value.

II. Physical Placement Optimization: Move the Ginger Scones to the display case nearest the espresso machine and Ouro Brasileiro shots. This creates a visual association, leveraging the high Lift (8x to 10x), which suggests a behavioral preference for the pairing.

III. Marketing Promotion: Create a targeted deal with the Ouro Brasileiro Shot + Ginger Scone, and specifically promote it during the strong pattern hours (Morning and Afternoon), reinforcing the observed co-occurrence.

Conclusion and Future Research Directions

Conclusion

The research successfully utilized time-segmented Apriori analysis to identify a strong and reliable item co-occurrence pattern: {Ouro Brasileiro shot} → {Ginger Scone}. This finding directly confirms the research question, providing a mechanism for dynamic inventory control. The implementation of the 10:7 stocking ratio (Table 2) during peak hours, combined with a significant reduction in Ginger Scone inventory in the evening when the pattern is absent, enables the café to significantly reduce surplus food while simultaneously optimizing stock levels to meet predictable demand. The overall effect is reduced surplus and an increased amount of savings through highly reliable upselling opportunities.

Future Research Directions

The next step is to implement the 10:7 stocking ratio for two weeks and track the sales and surplus reduction for Ginger Scones, particularly during the Evening segment. This provides empirical validation of the model's financial and environmental impact. Furthermore, to uncover "hidden gems" and moderate patterns for less frequent promotional use, future analysis should re-run the Apriori algorithm with lower thresholds, e.g., 30% confidence and 1% support. Then, to ensure the inventory strategy remains accurate and responsive, the analysis should be reassessed quarterly to monitor for shifts in consumer behavior due to seasonality, winter vs. summer menus, or menu changes.

References

Mercier, V. (2025, April 14). *How to reduce food waste in restaurants.* EHL Insights. https://hospitalityinsights.ehl.edu/food-waste-in-restaurants

Muhammad Abubakar, M. (2015). Association rule mining: market basket analysis of a grocery store. *International Journal of Science, Technology & Management*, *4*(5), 173–180.

ReFED. (2025, February). *From surplus to solutions: 2025 ReFED U.S. food waste report.* ReFED. https://refed.org/downloads/refed-us-food-waste-report-2025.pdf

Sigala, E. G., Gerwin, P., Chroni, C., Abeliotis, K., Strotmann, C., & Lasaridi, K. (2025). Reducing food waste in the HORECA sector using AI-based waste-tracking devices. *Waste Management*, *198*, 77–86. https://doi.org/10.1016/j.wasman.2025.02.044

Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). *Introduction to data mining* (2nd ed.). Pearson.

Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2024). *Understanding of data*. https://drive.google.com/file/d/1qM6_5vqBa5jpYHpecGBmn4PZfntR1p4q/view