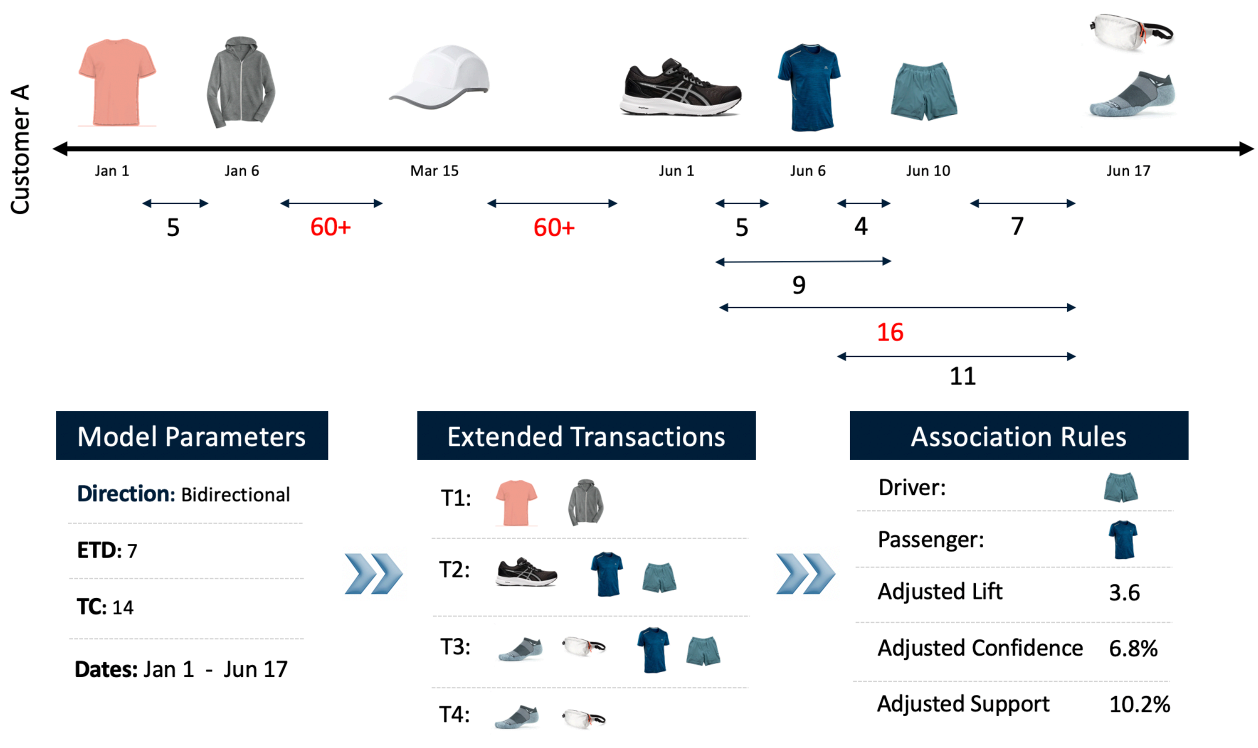
A MODERN-DAY ASSOCIATION MODEL

Steven Baez



SUMMARY

Steven's Extended Transaction and Bidirectional Model is a sophisticated association-based model designed to capture more complex consumer behaviors and provide greater insights into product pairings than traditional basket analysis models. It considers the buying behavior of consumers who purchase across platforms, such as online and retail stores, as well as the constraints that consumers may face, such as product availability or the need to make an afterthought purchase. The model also captures the behavior of consumers who purchase a single item that acts as a gateway to subsequent purchases.

At the core of Steven's Extended Transaction and Bidirectional Model is a bidirectional approach that analyzes transactions in both directions, from past to future and from future to past, allowing the model to consider both forward and backward relationships between items. This approach allows the model to identify both direct and indirect associations between items, providing a more comprehensive understanding of consumer behavior.

To build the model, stakeholders must set adjustable parameters based on their analysis, including Extended Transaction Days (ETD) and Transaction Continuity (TC) for the Extended Transaction Model, and Directional Weight for the Bidirectional Model. By adjusting these parameters, stakeholders can fine-tune the model to better capture the complexities of their specific market and consumer behavior.

Overall, Steven's Extended Transaction and Bidirectional Model provides a powerful tool for companies to understand their customers better and make data-driven decisions about product development, marketing strategies, and business operations. By analyzing complex buying behaviors across various platforms and customer segments, the model can help companies optimize their product offerings and marketing strategies to maximize sales and profitability.

THE BUSINESS NEED FOR ASSOCIATION MODELS

Consumer insights are becoming more important than ever for companies because they provide valuable information about the preferences, behaviors, and needs of their target audience. These insights help companies understand their customers better and make data-driven decisions about product development, marketing strategies, and business operations. It's also crucial for companies to understand how consumers purchase through various platforms, whether it's digital or brick-and-mortar, and adapt to the ever-changing ways to engage with consumers. However, with the abundance of point-of-sale data, companies often neglect how buying behavior across platforms can be used to extract greater product relationships.

To better understand the limitations of commonly used association-based models, let's start with a high-level overview of one widely used model - the market basket analysis model.

MARKET BASKET ANALYSIS MODEL – OVERVIEW

A market basket analysis model is a data analysis technique that examines the relationships between items that customers purchase together. It is also known as association analysis or affinity analysis.

The market basket analysis model uses transaction data from point of sale (POS) systems to identify which items are frequently purchased together by customers. The goal is to find patterns in customer behavior that can be used to inform marketing strategies, such as product placement, cross-selling, and targeted promotions.

The output of a market basket analysis model is typically a set of rules that describe the relationships between items. These rules are often expressed in the form of: "If a customer buys item X, they are likely to also buy item Y." These rules can be used to generate recommendations for customers, optimize store layout, and improve inventory management.

In practice, a market basket analysis model may be used by retailers to increase sales by promoting complementary items together, identifying popular product bundles, or optimizing store layouts. By understanding customer behavior and preferences, retailers can make data-driven decisions to improve the shopping experience for their customers and increase profitability for their business.

LIMITATIONS OF BASKET ANALYSIS MODELS

Although Market Basket Analysis models are widely used by consumer packaged-goods companies, they have their limitations. The primary limitation is that they extract product relationships by analyzing items purchased within the same transaction. This means that transactions containing only a single product - which account for approximately 60% of all transactions in consumer-packaged goods (CPG) - are typically discarded since they provide no insight into product relationships. This limitation can make it difficult to extract meaningful insights from transaction data and can impact the accuracy of product recommendations and other data-driven decisions.

Additionally, due to the model’s reliance on individual transactions, it can often neglect modern-day buying behaviors of consumers. Many consumers now shop across various sites and platforms in search of the best deals, which can result in transactions that are spread out across different platforms and days. As a result, the data used by Market Basket Analysis models may not capture the full picture of consumer behavior, making it difficult to accurately identify product relationships and associations that span across multiple transactions. This limitation underscores the importance of developing highly customized association-based models to better capture the complexities of modern consumer behavior.

To develop a modern-day Market Basket Analysis model, it's essential to build a model that can consider not only buying behaviors across various platforms and customer segments but also the time elapsed between individual transactions.

To illustrate the importance of accounting for lapsed time between transactions, let’s take a look at a specific example following one customers’ shopping behavior.

Suppose member ABC123 visits your company's local retail store with the intention of buying a new running outfit for the upcoming race season, including new running shoes, dry fit socks, dry fit running shorts, and a lightweight running shirt. However, the local store only carries 2 of the 4 items in ABC123's size. As a result, following day member ABC123 purchases the remaining items from your company's e-commerce site.

If a traditional market basket analysis modeling approach is used, your company may only consider one of the two transactions to generate insights into product relationships, likely the online transaction, or they may treat the two transactions as separate events, two separate transactions. However, either approach would yield little to no meaningful insight in product pairings.

STEVEN'S EXTENDED TRANSACTIONS - BASKET ANALYSIS MODEL

Steven's Extended Transactions - Basket Analysis model is designed to capture more complex consumer behaviors and provide greater insights into product pairings than traditional basket analysis models. It considers the buying behavior of consumers who purchase across platforms, such as online and retail stores. It also considers the constraints that consumers may face, such as product availability or the need to make a subsequent purchase.

Additionally, the model captures the behavior of consumers who purchase a single item that acts as a gateway to subsequent purchases. For example, a customer who buys running shorts and is satisfied with their purchase may soon after purchase a running shirt and headband in the days following their initial purchase. By analyzing consumer behaviors and grouping customer transactions based on predefined rules, companies can use single product transactions to identify frequently purchased products that would otherwise be disregarded. This information can help optimize product offerings and marketing strategies, ultimately maximizing sales and profitability.

At the core of Steven's Extended Transactions - Basket Analysis model are two adjustable parameters that stakeholders can modify based on their analysis: Extended Transaction Days (ETD) and Transaction Continuity (TC).

Here are high-level definitions of their intended use, but I will provide an example below to further illustrate their importance.

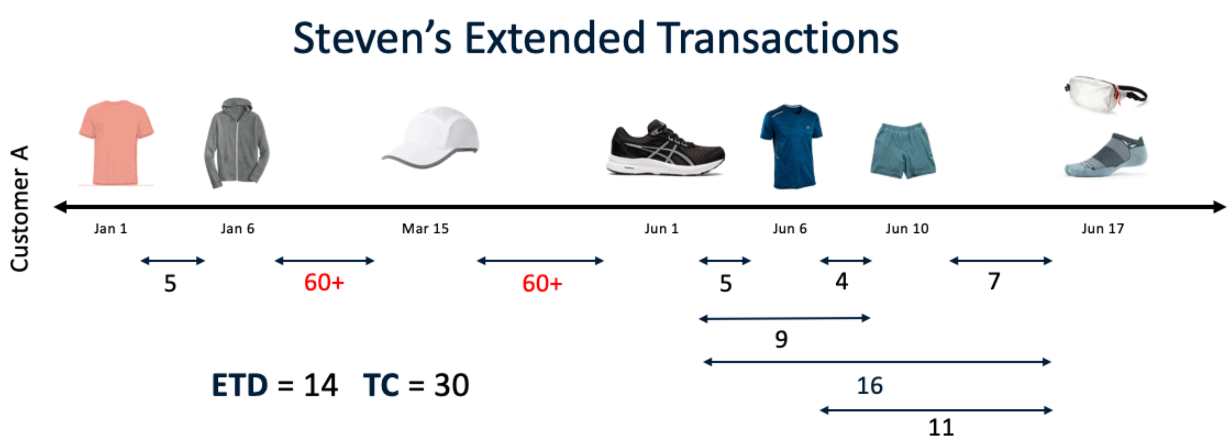
Extended Transaction Days (ETD): Maximum lapsed days between member transactions.

Transaction Continuity (TC): The number of allowable lapsed days from the first ETD transaction to the last.

By examining Customer A's transactions spanning six-months, as depicted below, we can demonstrate how ETD and TC can provide greater insights into product pairings compared to a conventional basket analysis.

A traditional market basket analysis models would limit its focus to Customer A's June 17th transaction, which involved the purchase of socks and a fanny pack, as depicted below. This is because all other transactions only comprise of single items, and a traditional market basket analysis model can only extract product pairings from multi-item transactions.

Steven's Extended Transactions - Basket Analysis model, in contrast, uses ETD and TC values to group transactions and support different use cases. When applied to Customer A's six-month transaction history with ETD set to 14 and TC set to 30, Steven's model generates extended transactions that provide more pertinent insights into product relationships. It accomplishes this by considering transactions that, for various reasons, had a brief interval between them.



UNDERSTANDING ETD AND TC

Extended Transactions are designed to accommodate lapses in product availability and consumer buying behaviors that result in separate subsequent purchases. By setting an Extended Transaction value of 14, all transactions that took place within 14 days of one another are grouped into a single transaction, at a customer level, and the date of the first transaction in the series is used as the extended transaction date. In theory, if a customer were to purchase a single item every week for a year, they could have a single extended transaction encompassing the entire year. To define the permanent end date of an Extended Transaction in terms of lapsed days from the first transaction in the series, Transaction Continuity is introduced. For instance, setting a Transaction Continuity value of 30 would terminate the Extended Transaction 30 days after the first transaction in the series, regardless of whether additional transactions meet the ETD criteria (i.e., are within 14 days of the last transaction in the series). All subsequent transactions are then grouped based on the set ETD and TC values, resulting in the creation of additional transactions.

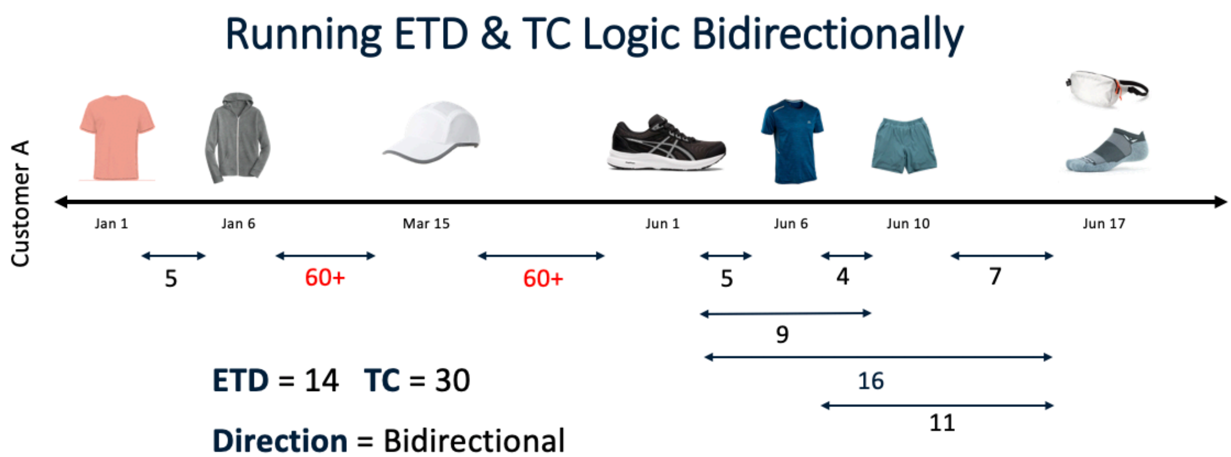
While the flexibility of setting Extended Transaction and Transaction Continuity values lies with the end user and their intended use case, Steven works with clients to develop additional machine learning models that can determine the appropriate values based on their unique customers' buying behavior.

By applying Steven's Extended Transactions - Basket Analysis model to Customer A's six-month transaction history, we can observe two extended transactions and one traditional transaction. The extended transactions allow more transactions to be passed to the statistical modeling component of a basket analysis model, which computes metrics such as Lift, Confidence, and Support. This results in variations of each metric.

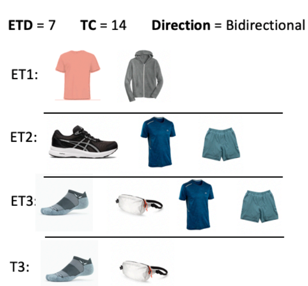
STEVEN'S BIDIRECTIONAL EXTENDED TRANSACTIONS MODEL

Steven has further developed his original Extended Transactions model to run in both directions, providing even more detailed product insights. In situations where the sequence of product purchases is less relevant, or transaction counts are low, resulting in inflated Lift, Confidence, and Support values, a bidirectional approach is more appropriate. By running Steven's Extended Transactions – Basket Analysis model bidirectionally, more transactions that meet the defined ETD and TC criteria can be generated than Steven's Extended Transaction model.

Let's consider an example in which we examine Customer A's six-month transaction history, using ETD of 7, TC of 14, and a bidirectional model direction.



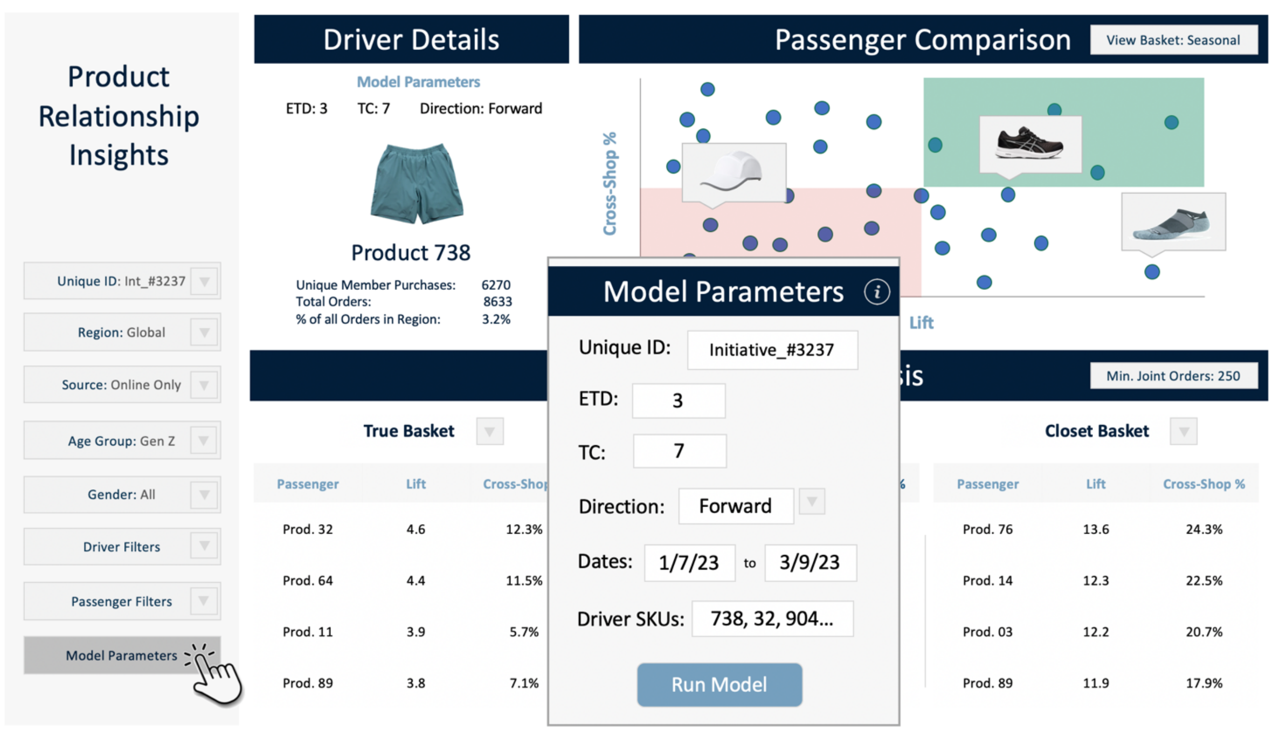
By utilizing Steven's Bidirectional Extended Transactions model and setting ETD to 7 and TC to 14, we are able to create three extended transactions from Customer A's six-month purchasing history. Unlike a traditional Market Basket Analysis Model that would only return a single transaction, Steven's model generated four transactions by uncovering buying patterns and relationships hidden amongst both multi and single item transactions. When it comes to product relationships, particularly association models, having more relevant transactions is crucial for extracting meaningful statistical insights about products.



This highlights the increased value of utilizing an association-based model that can pass through additional transactions to gain a better understanding of product relationships. Moreover, users have the flexibility to adjust the Direction Parameter to either "Forward", "Backward", or "Bidirectional", depending on the importance of the transaction sequence in comprehending relationships.

Using Steven's bidirectional approach and ETD/TC values, users are able to identify a previously undiscovered relationship between athletic tops and athletic bottoms through the creation of transaction two and three (shown in image above). This insight would have been missed using a traditional Market Basket Analysis, highlighting the importance of incorporating ETD and TC values to accurately reflect consumer buying patterns. The combination of ETD, TC, and bidirectional analysis in Steven's Extended Transactions – Basket Analysis model provides a more comprehensive understanding of product relationships and can generate more valuable insights for businesses in the consumer packaged-goods industry.

ENABLING USERS TO INTERACT WITH MODEL PARAMETERS IN REALTIME



Steven has worked with clients to build and implement Product Relationship Insight dashboards, which allow end-users to run Steven's Bidirectional Extended Transactions in real-time through sophisticated interfaces. You can contact Steven to request a demonstration of the dashboard and its underlying Bidirectional Extended Transactions.

You may also want to consider the following association models that are most suitable for your business needs:

1. Apriori Algorithm: A widely used algorithm for association rule mining, which generates frequent itemsets by iteratively reducing support threshold, and then derives association rules from the frequent itemsets.
2. FP-Growth Algorithm: An efficient algorithm for mining frequent itemsets, which uses a compact data structure called FP-tree to store frequent itemsets.
3. ECLAT Algorithm: A vertical data format-based algorithm for frequent itemset mining, which projects the dataset into subsets and searches for frequent itemsets in each subset.
4. Market Basket Analysis: A traditional approach to association rule mining, which analyzes the purchasing behavior of customers in a retail environment to uncover relationships between products.
5. Steven's Bidirectional Extended Transactions - Basket Analysis model: is an advanced data analysis technique that builds upon the traditional market basket analysis approach. By incorporating two additional parameters, Extended Transaction Duration (ETD) and Transaction Continuity (TC), this model groups multiple transactions based on their proximity in time and the duration of the transactional period. Steven's bidirectional approach also considers both forward and backward product sequences in order to generate more transactions and reduce the possibility of inflated support, confidence, and lift scores.
6. Collaborative Filtering: A technique that analyzes user behavior to recommend items, often used in recommendation systems for online marketplaces and streaming services.
7. Sequence Mining: A technique that identifies frequent sequences or patterns in a dataset, often used to understand the behavior of customers or users over time.
8. Association Rule Hiding: A technique used to protect the privacy of sensitive association rules in datasets by hiding or removing them while preserving the overall accuracy of the model.