

In the business intelligence world, “market basket analysis” helps retailers better understand – and ultimately serve – their users by predicting their purchasing behaviors. In this blog post, we’ll explain how market basket analysis works and what it takes to deploy a market analysis project.

MBA aims to find relationships and establish patterns across purchases. The relationship is modeled in the form of a conditional algorithm:

IF {sandwich, cookies} THEN {drink}

In the shorthand notation, which translates to “the items on the right are likely to be ordered with the items on the left:”

$$\{A_i\} \rightarrow \{C_i\}$$

A collection of items purchased by a customer is an **itemset**. The set of items on the left-hand side (sandwich, cookies in the example above) is the **antecedent** of the rule, while the one to the right (drink) is the **consequent**. The probability that the antecedent event will occur, i.e., a customer will buy a sandwich and cookies, is the **support** of the rule. That simply refers to the relative frequency that an itemset appears in transactions. In a QSR, the support of an item or item combination helps to identify keystone products. Hence, if a sandwich and cookies have high support, then they can be priced to attract people to the store.

The probability that a customer will purchase a drink on the condition of purchasing a sandwich and cookies is referred to as the **confidence** of the rule. Confidence can be used for product placement strategy and increasing profitability. Placing high margin items near associated high confidence (driver) items can increase the overall margin on purchases.



According to market basket analysis, a customer that orders this meal would be more likely to order a drink.

The **lift** of the rule is the ratio of the support of the left-hand side of the rule (sandwich, cookies) co-occurring with the right-hand side (drink), divided by the probability that the left-hand side and right-hand side co-occur if the two are independent.

- ✓ A lift greater than 1 suggests that the presence of the antecedent increases the chances that the consequent will occur in a given transaction
- ✓ Lift below 1 indicates that purchasing the antecedent reduces the chances of purchasing the consequent in the same transaction. Note: This could indicate that the items are seen by customers as alternatives to each other
- ✓ When the lift is 1, then purchasing the antecedent makes no difference on the chances of purchasing the consequent



Market basket analysts search for rules with lift that are greater than 1 backed with high confidence values and often, high support.

The **Apriori algorithm** is a commonly-applied technique in computational statistics that identifies itemsets that occur with a support greater than a pre-defined value (frequency) and calculates the confidence of all possible rules based on those itemsets.

Market Basket Analysis Example

The Apriori algorithm is implemented in the **arules package**, which can be installed and run in **R**. Data is loaded into the engine in the following format:

Transaction ID	Item Name
1	Sandwich
1	Cookies
1	Milk
1	Egg
2	Burger
2	Pizza
2	Salad
2	Egg
2	Bottled Water
2	Chicken
3	Sandwich
3	Egg
3	Bottled Water

The first column is the order/transaction number and the second is the item name or, more often, the item ID. The next steps usually involve aggregating each transaction across records into a single record as an array and converting the dataset to an R transaction. The result of that aggregation is as shown below:

Transaction ID	Items List
1	Cookies, Egg, Milk, Sandwich
2	Bottled Water, Burger, Chicken, Egg, Pizza, Salad
3	Beacon, Bottled Water, Egg, Sandwich, Yogurt
4	Burger, Pie, Pizza, Salad, Soda
5	Burger, Ice Cream, Pie, Pizza, Salad, Soda
6	Chocolate Shake, Cookies, Egg, Milk, Sandwich
7	Beacon, Chocolate Shake, Cookies, Milk, Yogurt
8	Bottled Water, Burger, Chicken, Chocolate Shake, Egg, Pie, Pizza, Salad, Soda
9	Beacon, Bottled Water, Egg, Milk, Pizza, Salad, Yogurt
10	Chocolate Shake, Cookies, Egg, Milk, Sandwich
11	Beacon, Burger, Salad
12	Cookies, Egg, Milk, Sandwich, Yogurt
13	Beacon, Bottled Water, Egg, Pie, Pizza, Sandwich
14	Cookies, Egg, Milk, Sandwich
15	Bottled Water, Burger, Chicken, Egg, Pie, Pizza, Salad

Finally, the Apriori logic is applied to the transactions with the following **resultset**:

LHS	RHS	rules	support	confidence	lift
Ice Cream	Soda	{Ice Cream} => {Soda}	0.07	1.00	5.00
Soda	Ice Cream	{Soda} => {Ice Cream}	0.07	0.33	5.00
Ice Cream	Pie	{Ice Cream} => {Pie}	0.07	1.00	3.00
Pie	Ice Cream	{Pie} => {Ice Cream}	0.07	0.20	3.00
Ice Cream	Burger	{Ice Cream} => {Burger}	0.07	1.00	2.50
Burger	Ice Cream	{Burger} => {Ice Cream}	0.07	0.17	2.50
Ice Cream	Salad	{Ice Cream} => {Salad}	0.07	1.00	2.14
Salad	Ice Cream	{Salad} => {Ice Cream}	0.07	0.14	2.14
Ice Cream	Pizza	{Ice Cream} => {Pizza}	0.07	1.00	2.14
Pizza	Ice Cream	{Pizza} => {Ice Cream}	0.07	0.14	2.14
Soda	Chicken	{Soda} => {Chicken}	0.07	0.33	1.67
Chicken	Soda	{Chicken} => {Soda}	0.07	0.33	1.67
Soda	Chocolate Shake	{Soda} => {Chocolate Shake}	0.07	0.33	1.25
Chocolate Shake	Soda	{Chocolate Shake} => {Soda}	0.07	0.25	1.25
Soda	Pie	{Soda} => {Pie}	0.20	1.00	3.00
Pie	Soda	{Pie} => {Soda}	0.20	0.60	3.00
Soda	Burger	{Soda} => {Burger}	0.20	1.00	2.50
Burger	Soda	{Burger} => {Soda}	0.20	0.50	2.50
Soda	Bottled Water	{Soda} => {Bottled Water}	0.07	0.33	0.83
Bottled Water	Soda	{Bottled Water} => {Soda}	0.07	0.17	0.83
Soda	Salad	{Soda} => {Salad}	0.20	1.00	2.14
Salad	Soda	{Salad} => {Soda}	0.20	0.43	2.14
Soda	Pizza	{Soda} => {Pizza}	0.20	1.00	2.14

Overall, we see a number of strong consequent (RHS) combinations with soda, which could indicate that soda is a keystone product category. A manager could decide to keep the price and margins low on soda to drive sales volume. Bottled water is the exception,



but we can assume that those would be considered alternatives to each other.

Perhaps a more interesting result is that all of the rules with ice cream show significant lift with a confidence of 1. (Associations in the other direction aren't nearly as likely.) This could lead us to further promote ice cream with the expectation that customers will purchase other items at the same time.

Other Uses for Market Basket Analysis

In addition its popularity as a retailer's technique, MBA is applicable in many other areas:

- ✓ Manufacturing: predictive analysis of equipment failure
- ✓ Pharmaceutical/Bioinformatics: discovery of co-occurrence relationships among diagnosis and pharmaceutical active ingredients prescribed to different patient groups
- ✓ Financial/Criminology: fraud detection based on credit card usage data
- ✓ Customer Behavior: associating purchases with demographic and socio-economic data

More and more organizations are discovering ways of using market basket analysis to gain useful insights into associations and hidden relationships. As industry leaders continue to explore the technique's value, a predictive version of market basket analysis is

