

Digging Deeper – Mining Patterns Beneath the Surface with Association Rules

By **Ben Cortese** (Statistician)

Traditional modeling techniques focus on high level trends and identify general patterns that reflect a majority of customers. The less obvious findings within smaller groups can be hidden from view. These smaller groups may be the key to moving your business forward, particularly in situations that include:

- Mature brands and products where a general overview of the business is well understood
- Established businesses that are looking to grow their shares
- Similarities between a business and its competitors

All customers have an important voice, even those that do not make up a large portion of the population. One solution to better understand these groups and how to influence them is association rule mining, a technique that identifies a wide variety of relationships buried within the data. This method digs deeper by filling in the gaps through the identification of sub segments with similar makeup, whether that be attitudes, demographics, or any other survey data, and an associated action each group is likely to take.

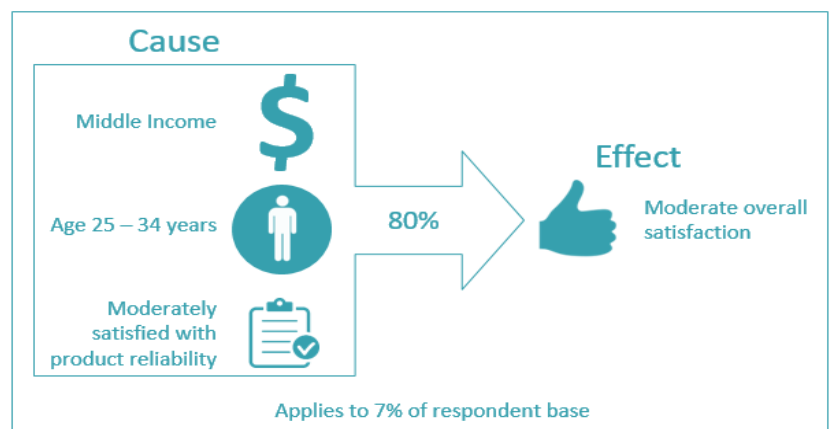
In most cases, association rule mining generates many thousands of rules. An association rule is a set of attributes that commonly occur together, split into two types – the cause and the effect. The effect is a single attribute that is caused by the remaining attributes identified in the rule. Effects may include satisfaction, market share, purchasing behavior, lifetime value, likelihood to switch providers, and more.

The enormous variety of rules available opens the door for exploration of specific groups, attributes, attitudes, or any other hypotheses of interest by filtering the rules. Filtering greatly reduces the number of rules and is based on:

- The strength of the association of the cause and effect
- The proportion of the respondent base the rule applies to
- The attributes in the cause and/or effect

The ability to focus on rules that apply to a small percentage of the respondent base can help uncover less obvious patterns that may have been overlooked in other analyses.

As an example, consider Acme Co. – a company interested in expanding its market share. A rule mined from survey data that aligns with 7% of the respondent base is displayed in the graphic and can be interpreted as follows. Eighty percent of respondents that have the causal attributes in this rule also had moderate overall satisfaction ratings (the other 20% rated overall



satisfaction either high or low). This implies that focusing on product reliability may improve overall satisfaction for this group of respondents and in turn, lead to a larger market share.

Now, suppose the same company would like to better understand how to position themselves with respect to a leading competitor. The following table includes a snippet of the rules that are mined to differentiate Acme Co. from its competitor.

Female, High innovation rating , High licensing rating	High likelihood to recommend Acme Co.
Middle income, High knowledgeable rating	High likelihood to recommend Acme Co.
Age 45 – 54 years, High innovation rating , High knowledgeable rating	High likelihood to recommend Acme Co.
Female, High account team rating, High value rating	High likelihood to recommend Competitor
Middle Income, High value rating , High knowledgeable rating	High likelihood to recommend Competitor

While both companies are viewed as highly knowledgeable by the middle income segment, **high innovation** ratings are prevalent only for Acme Co. Their competitor is viewed by the same segments as having **high value**. This differentiation in both the middle income and female segments indicates the strengths of both Acme Co. and its competitor. In order to obtain a larger share of the female and/or middle income segments, Acme Co. should focus on improving value of its offerings while maintaining their innovative approach.

Association rules provide deeper insights of the details of your business that are often times hidden by traditional modeling techniques. This technique can help you better understand where to focus your resources to have an impact on smaller segments of your customer base.

About the Author

Ben Cortese has been with KS&R for nearly two years, working with a diverse client group on advanced analytics projects. As an expert R programmer, his focus is on exploratory analysis, Bayesian network modeling, key driver analysis, discrete choice and predictive modeling.



Education: PhD (Statistics), Syracuse University

[Return To Thought Leadership >](#)