Association Between Products

By: Usman Iftikhar

Electronidex, a start-up electronics online retailer, provided it’s 30-day transactional data for all the electronics currently being sold. The online retailer’s transactional data consists of 9,835 online transactions from 125 products that are broken down into 17 product types. In determining purchasing patterns from the given transactional data, Market Basket Analysis is utilized. Market Basket Analysis is an immediate extension of association rules (made up of two or more items) with business outcome. Association rules are patterns discovered between items and their associated transactions.

Examining the data shows iMac to be the most frequently purchased item that showed up in 2,519 transactions. List of the top five most frequently purchased items can be found in Table 1. The transactional data also shows that as the number of items (itemset) increase in a transaction, the frequency of their occurrence decreases, as illustrated in Figure 1. This intuitively makes sense as the frequency of a single item purchased is likely higher than two items purchased together, and the frequency of two items purchased together is likely higher than three items purchased together, and the pattern continues. The frequency of a single item purchased was 2,163 transactions and the frequency of two items purchased together was 1,647, and the frequency continues to decline as the itemset grows.



Table . Top five most frequently occurring items in transactions.

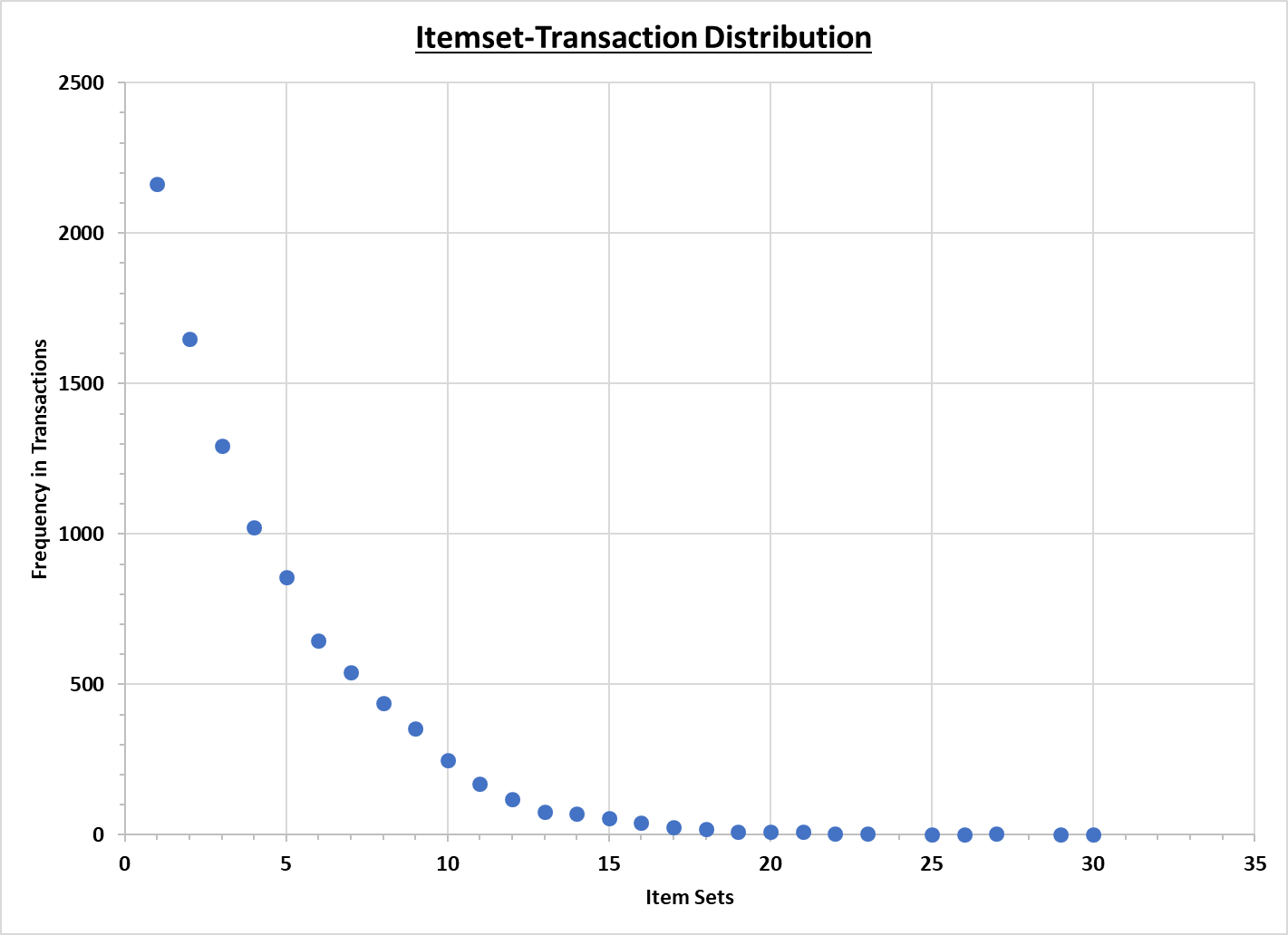


Figure . Itemset-Transaction relationship.

Further analysis of Electronidex’s transactional data revealed association between different products that are purchased together in the same transactions. Terminology used to investigate the transactional data and generate insights, is provided below.

* **Support** – how frequently an item occurs in the transactional data set (popularity of an itemset).
* **Confidence** – proportion of transactions where the presence of item (or a set of items) results in the presence of another set of items.
* **Lift** – likelihood of itemset A being purchased when itemset B is purchased while considering the popularity of A. A lift value of greater than 1 implies significant association, and in general, the larger the lift ratio, the more significant the association.

Figure 2 illustrates the top 10 association rules generated in accordance with the above terminology, sorted by the highest to lowest lift. The size of the circle represents the lift and the color of the circle represent the support. For instance, Acer Aspire and HP Laptop products both point to the largest circle, and an arrow leading away from this circle is directed to ViewSonic Monitor. This signifies the most significant rule and is also shown in Table 2. Similarly, examining rule 2 from Table 2 that shows arrows coming from Acer Desktop and HP Laptop lead to a circle that is not as big nor as bright and an arrow leading from this circle to Dell Desktop. This association rule is not as strong as rule number 1. From the top 10 rules shown in the illustration and the table below, it is evident that ViewSonic Monitor appears most frequently with laptop purchases. Patterns such as this indicate sales initiatives of bundling monitors and laptops. Blackwell would benefit from this acquisition since it’s existing product lineup includes PCs and Laptops of similar brands. This acquisition can also lead to the opportunities of cross-selling different items.

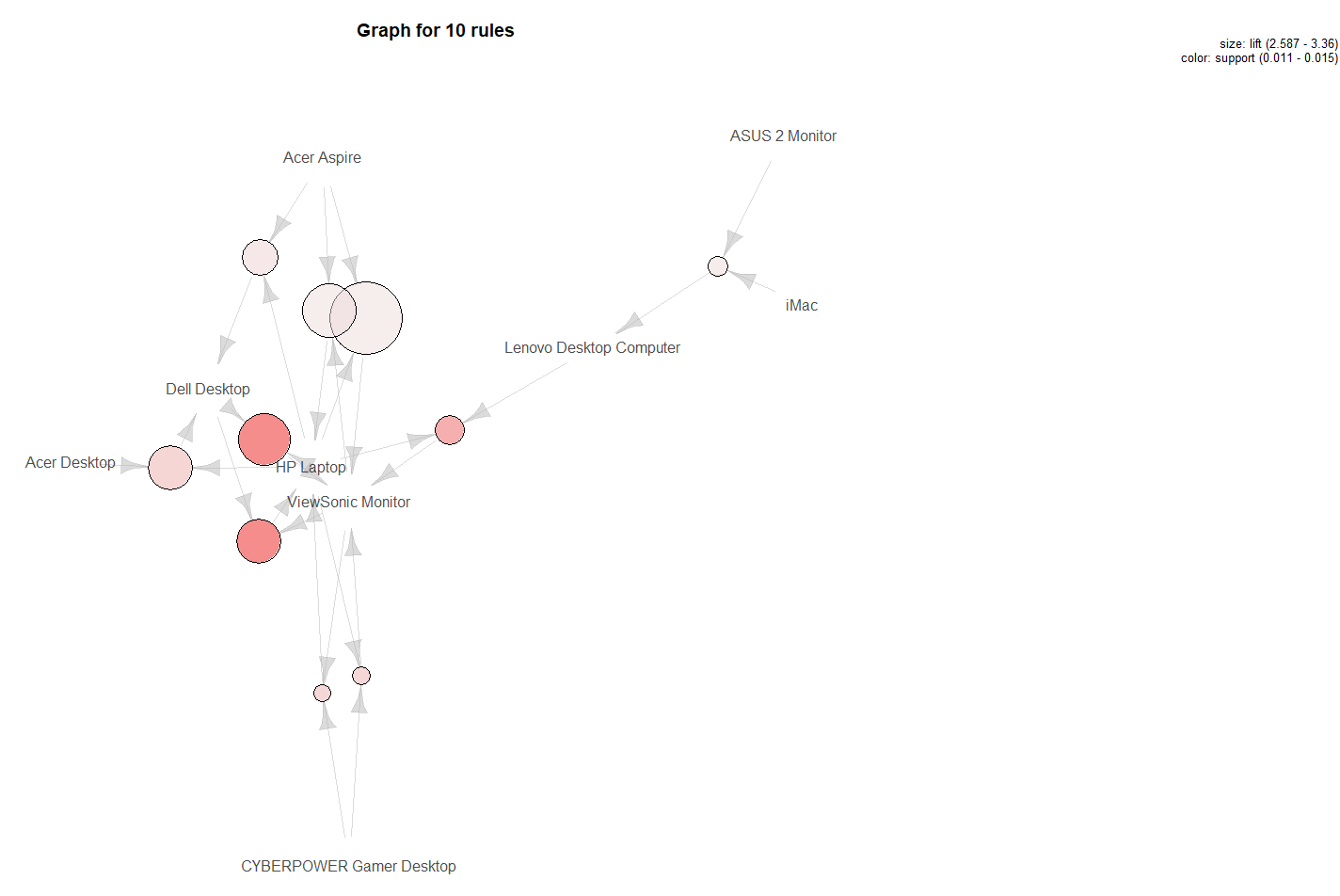


Figure . Illustration of top 10 rules association sorted by highest lift parameter.



Table . Top 5 associated rules sorted by lift parameter (table format of Figure 2 illustration).