Discover Associations Between Products: Market Basket Analysis

The objective of this analysis is to look into Electronidex’s clientele transactions and investigate how the purchase of certain items might lead to the client being more inclined to purchase an additional product based on his transaction history and other customers’ transactions. After going through my findings, I will discuss whether or not Blackwell should acquire Electronidex, and if so, what recommendations I have for Blackwell in order for the acquisition to be successful.

Electronidex provided me with a month’s worth of online transactions with the items corresponding to each transaction. This turned out to be a data set of 9,835 transactions with a possible 125 items offered by Electronidex. The items are all things electronics, ranging from hardware to software. One item per transaction was the most common type of transaction at 2,163 of the transactions being a single item purchase, or about 1 single item transaction for every 5 transactions made. The second most common number of items per transaction was 2 items at 1,647 of those transactions made. The third most common is 3 items per transaction at 1,294 transactions made with 3 items. The median number of number of items purchased is 3. This means that half of the transactions made were of less than 3 items and vice versa. However, the average number of items purchased per transaction came in at 4.383. This means that, even though half of the people purchased 3 or less items, the people responsible for transactions of 3 or more items, purchased a lot more than 3 item. The transaction with the most items contained 30 items. The top 5 items sold were iMacs, HP Laptops, CYBERPOWER Gamer Desktops, Apple Earpods and Apple MacBook Airs, Electronidex sold 2519, 1909, 1809, 1715 and 1530 respectively. More than 25% of all transactions contained an iMac as an item, or 1 in 4 transactions. Below is a frequency graph plotting the top 5 items most frequently found in any one transaction.

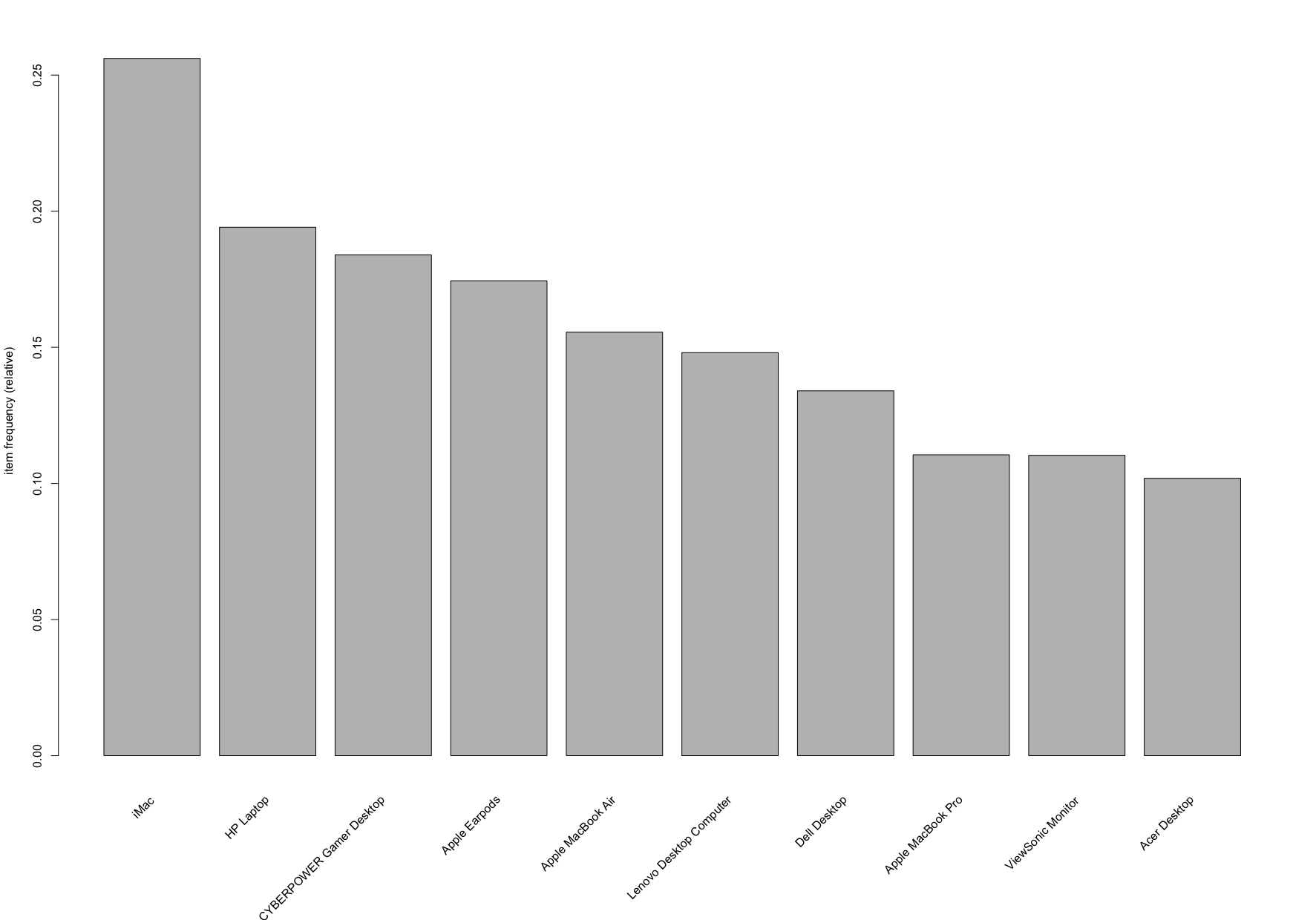


Figure : Top 10 Items Purchased Relative to Total Number of Transactions Made

Once I had gotten familiar with the data set I was working with, I began to look for correlations between items across transactions. I began by creating an image that sampled 125 transactions and marked the items purchased in each of the transactions.

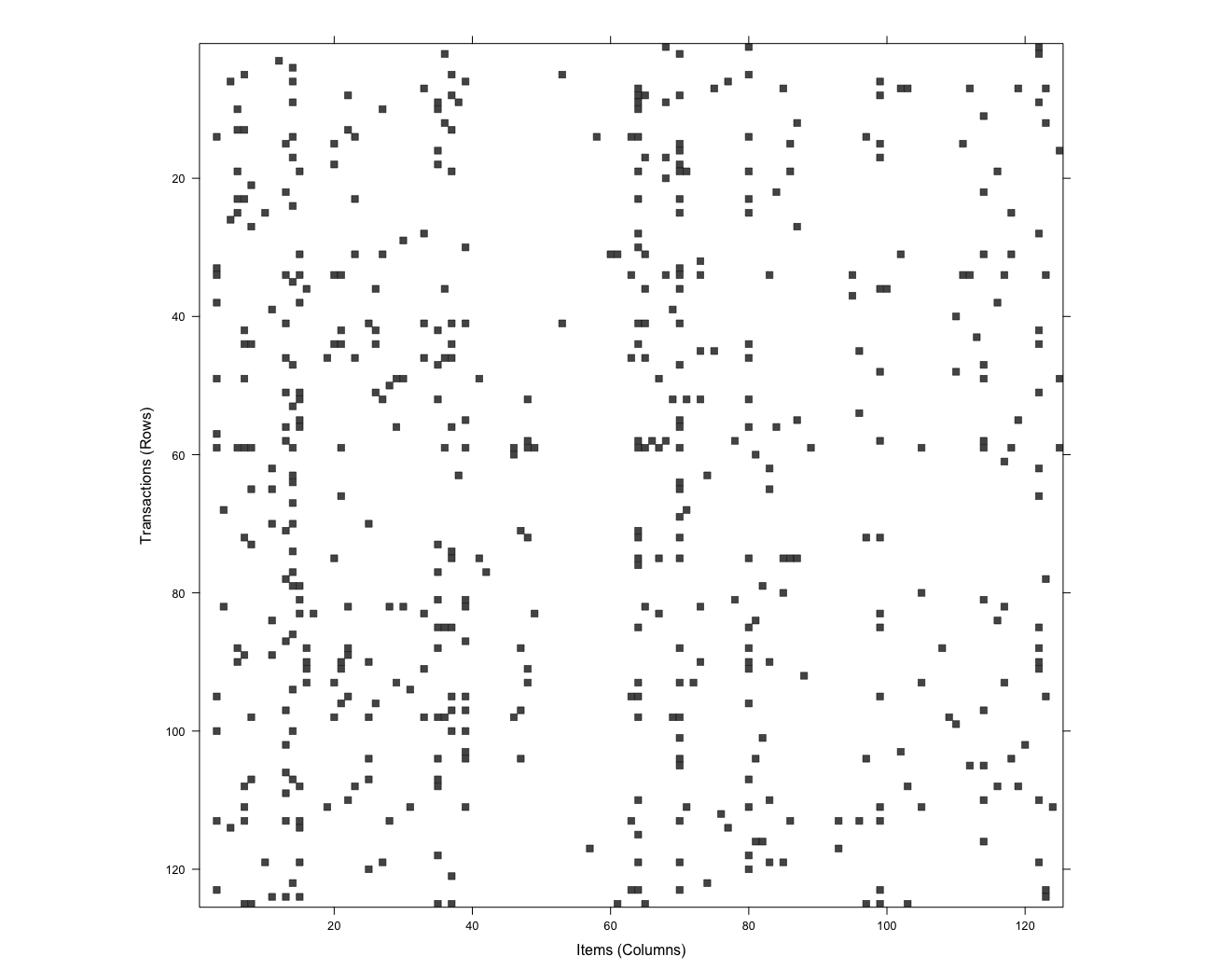


Figure : 125 Transactions Plotted vs the 125 Possible Items

Figure 2 plots 125 random transactions against all 125 items sold at Electronidex; the x axis contains the items, while the y axis contains the sampled transactions. Every black cell represents an item being purchased; you can check what items were purchased on any transaction by finding the desired transaction and going across horizontally to the right. It can be seen that items 50 to 60 are not commonly purchased, because most of the transactions have that item range in blank. On the other hand, you can see a potential correlation inside the red square. The transactions in the red squared, marked with the dashed lines have similar item transactions, but both are missing some items from each other’s transaction. This could help us figure out what items can be suggested or recommended to a customer based on another similar transaction made by another customer. These ideas are all educated guesses based on images, however, there is a way of quantitatively determining what products are commonly purchased with one another. We can then use that knowledge to more accurately recommend other items to clients. Using a market basket algorithm, Apriori Algorithm, I was able to develop 19 rules that could help the company suggest additional items to a client.

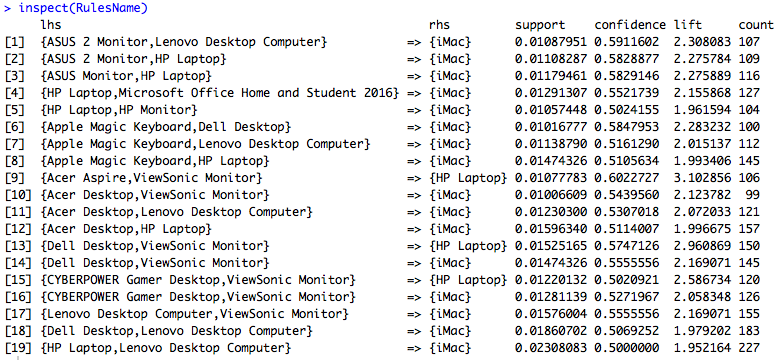


Figure : 19 Rules that Show Possible Correlations Between Items

In order to come up with these rules I had to take into account three parameters, the support value, confidence level and the lift. The support value determines the minimum amount of transactions in the data set that have to share the same combination of items in a transaction. I chose a support value of .01 of all transactions, which means at least about 99 of the 9,835 transactions must contain that set of items. Confidence level determines how many transactions contain the recommended item given we already have the other items in the item set we’re looking at relative to the total number of transactions; confidence level gives you the probability the client is going to purchase the recommended item given his transaction history. I chose a minimum confidence level of 50%. This means that there is a 50% chance that the customer will find interest in a recommended item and purchase it. Last but not least, lift tells us how important a certain rule is, and that the shared item set the items share, is not due to chance. A lift value of 1 indicates low importance, while anything close to 2 and above indicates a correlation not due to luck. The lowest lift value associated with one of the 19 rules is about 1.95; this indicates that even the lowest lift value is acceptable, and the rules are not due to chance. Rule 9 contained both the highest confidence level and lift value of 0.602 and 3.103 respectively. Rule 19 contained the highest support level of 0.023. All of the rules recommended either iMacs or HP Laptops, and is probably because iMacs and HP Laptops are the two most common items purchased. Even though all of these rules passed the minimum quantitative test, there is still three categories we can group them into. Insightful rules are rules that offer a new discovery or insight into a relationship between items that wouldn’t have been possible without the analysis. Irrelevant rules are rules that are obvious and don’t contribute to any further understanding of our customers. Lastly, unclear rules are rules that are not obvious, but they are rules that do not seem to make sense even after discovering them. I placed Rules 2,3,4,9,13,14,15,16 under insightful rules due to the fact that they make sense but probably would not have been discovered without the market basket analysis. Rules 4,8 fell under irrelevant rules due to the obvious statements they make. The rest of the rules fell under the unclear rules category due to their puzzling nature that even after being shown, I do not seem to understand.

In conclusion I would recommend Blackwell to acquire Electronidex given the large number of transactions made in a single month. I would suggest Blackwell follow the rules found through the market basket analysis and apply it to the future customers. This will increase the amount of sales by making the clientele aware of other products they would like to purchase on their next transaction. Also, I suggest that Blackwell focus its attention in recommending iMacs to every customer, since, according to the data, 1 out of every 4 customers will buy it. Below are extra visuals created along with a brief explanation for each.

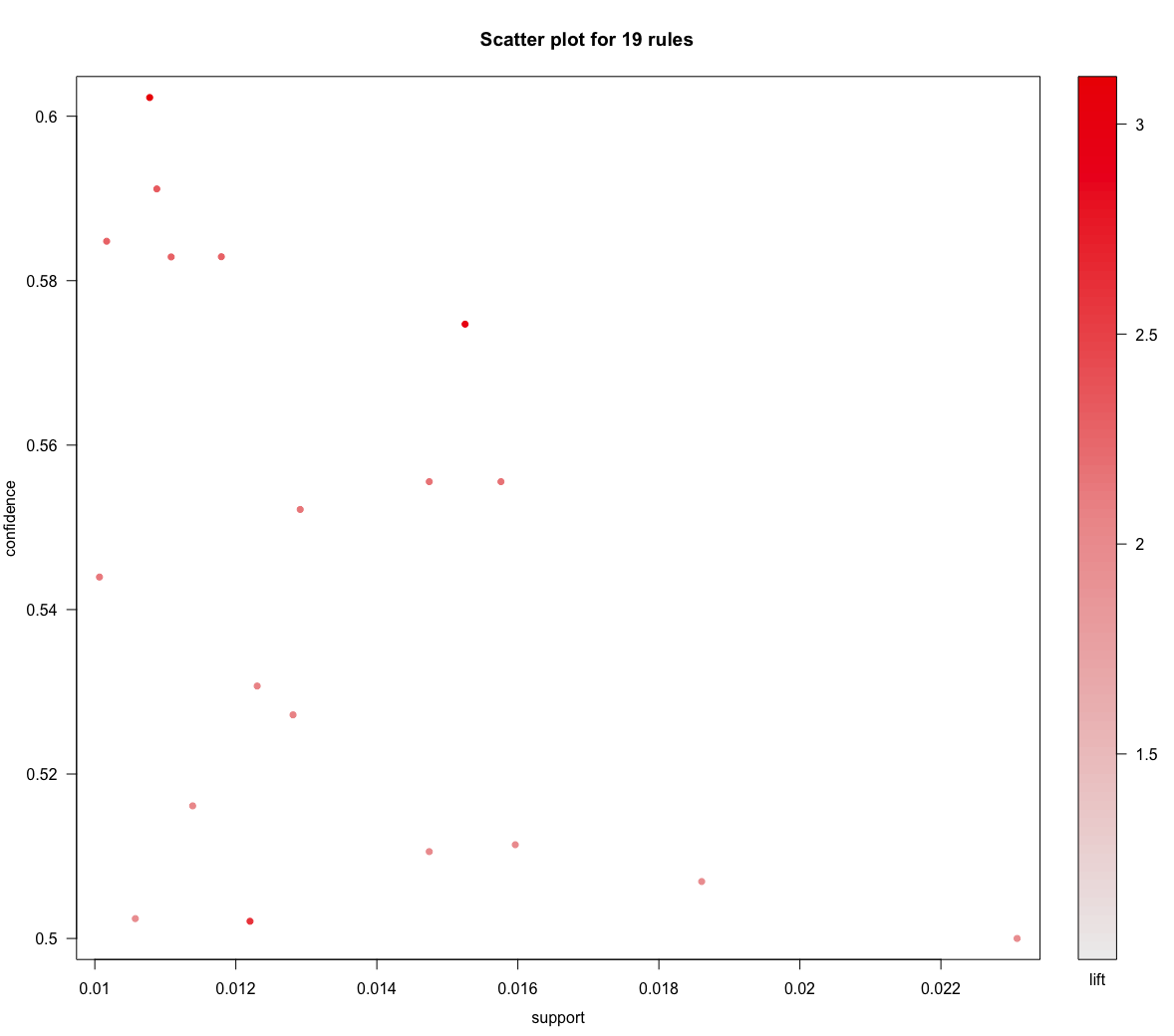


Figure : Scatter Plot of 19 Rules w/ Confidence Level, Support Value and Lift

From Figure 4 we can see that the rules with the highest support values have the lowest confidence levels and vice versa. Rule 13, enclosed in the blue square, seems to be the most balanced rule with a confidence level of 0.574, a support value of .015 and a high lift value of 2.961.

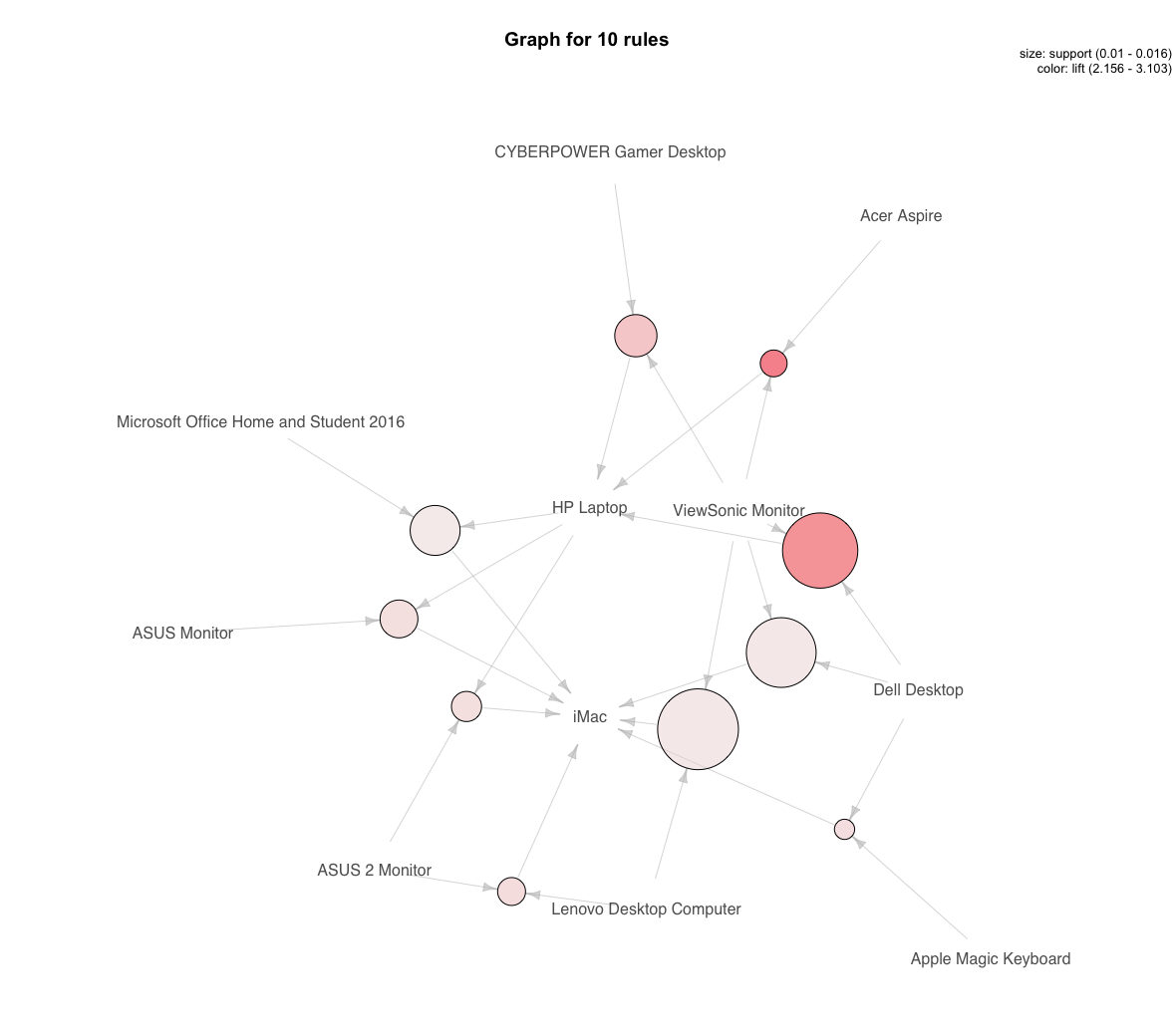


Figure : Plot of 10 Rules w/ Highest Lift Values

In Figure 5, the darker the dot the higher the rule’s lift value is, and the bigger the dot, the higher support value the rule has.