# Task 4: Discover Associations between Products

This informal report contains a market basket analysis performed on a dataset provided by Electronidex in order to answer the following questions:

* Are there any interesting patterns or item relationships within Electronidex's transactions?
* Would Blackwell benefit from selling any of Electronidex's items?
* In your opinion, should Blackwell acquire Electronidex?
* If Blackwell does acquire Electronidex, do you have any recommendations for Blackwell? (Ex: cross-selling items, sale promotions, should they remove items, etc.)

The dataset we have received contains 9835 transactions over 125 different objects. Below we can see the most frequently purchased items:

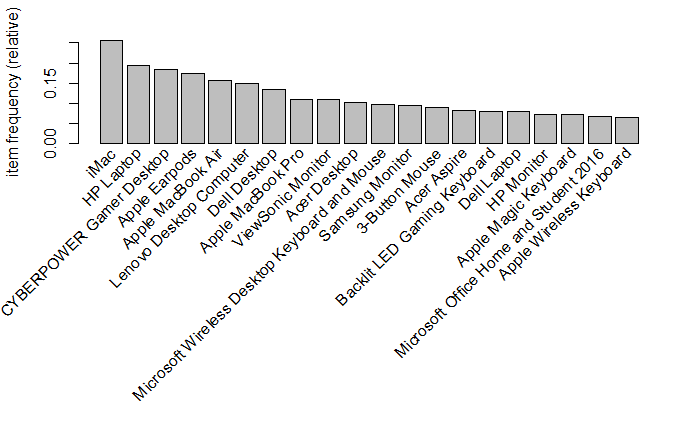


Figure 1 - Top 20 items purchased by frequency in relative terms

The density of the dataset is 0.035, meaning that a total of (9835 x 125 x 0.035 =) 43104 items have been purchased in total. And the most frequently purchased quantity of items per transaction:

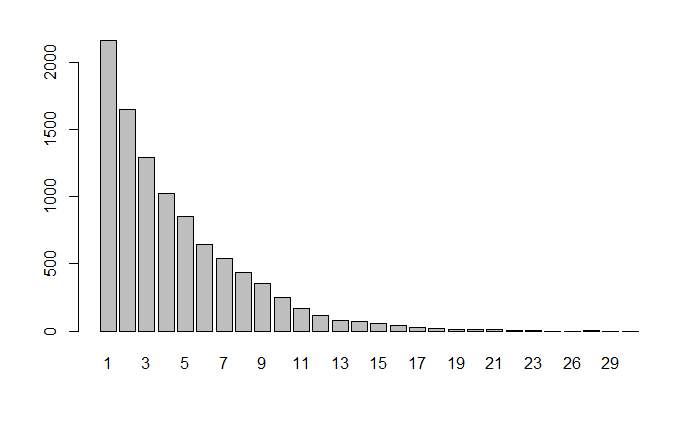


Figure 2 - Quantity of items purchased per transaction

As expected, this indicates that most customers buy a small number of items in each transaction. Additionally the size of purchases seems right skewed (as the mean is larger than the median an as can be seen below in the barplot of the table of transaction sizes).

## Apriori Model:

The Apriori algorithm assesses association rules using two types of measurements. The aim is to find strong rules, rules that measure high in both support and confidence. In order to catch these, I defined that rules cover at least 1% of the transactions and are correct at least 30% of the time. A support of 1% may seem low, but this shop offers a wide range of items and repetitions do not happen too often.

This level of support and confidence provides 152 rules (2 additional ones were redundant and removed). The rule with the best confidence is slightly larger than 60% with a support of 1.07% and the rule with the highest support has got 7.5%.

Out of these rules many were either irrelevant or unclear, only a few were actually insightful. Additionally many rules have been created with the “HP Laptop” and the “iMac” as the consequent (or right hand side of the transaction rule) because these were the most commonly purchased items.

Many rules may seem relevant, but in many cases this is just because they cover a lot of transactions. For instance, looking at the parallel coordinates graph below, it does not seem obvious that an individual who purchased a monitor and a desktop will clearly buy a laptop or another desktop (even though the apriori model suggests having a high confidence). I would say that the majority of clients who purchase these types of items together are not individuals like me but companies who require to purchase many items of the same type at once for their multiple employees.

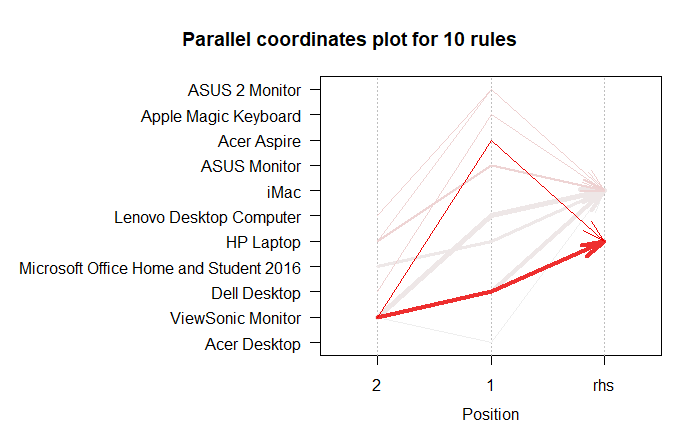


Figure 3 - Parallel coordinates for top 10 rules in terms of confidence.

I then decide to sort the rules according to the support and got the following results for the top 10:

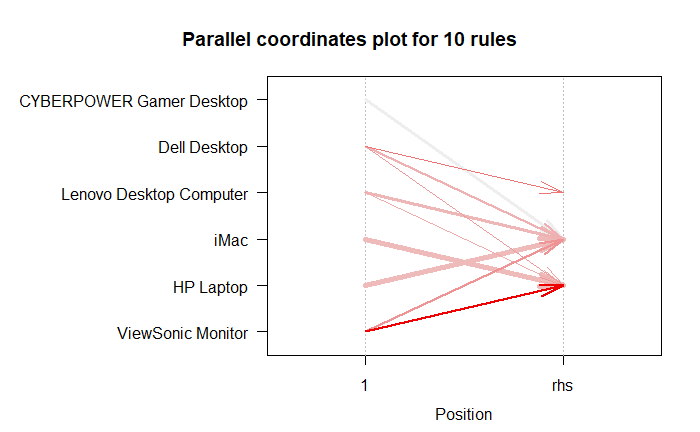


Figure 4 - Parallel coordinates for top 10 rules in terms of support.

There is a clear difference between the rules in figure 3 and 4. Sorting the rules according to support instead of confidence reduces the amount of items on the left hand side of the rule. I assume it is no coincidence that the most occurring rules, are also the less exclusive as well. Once again many of these rules contain items of the same type which make them harder to explain unless we are talking about an organization or company making the purchase. Below, follows an alternative visualization method of the same data shown above done by grouping the data.

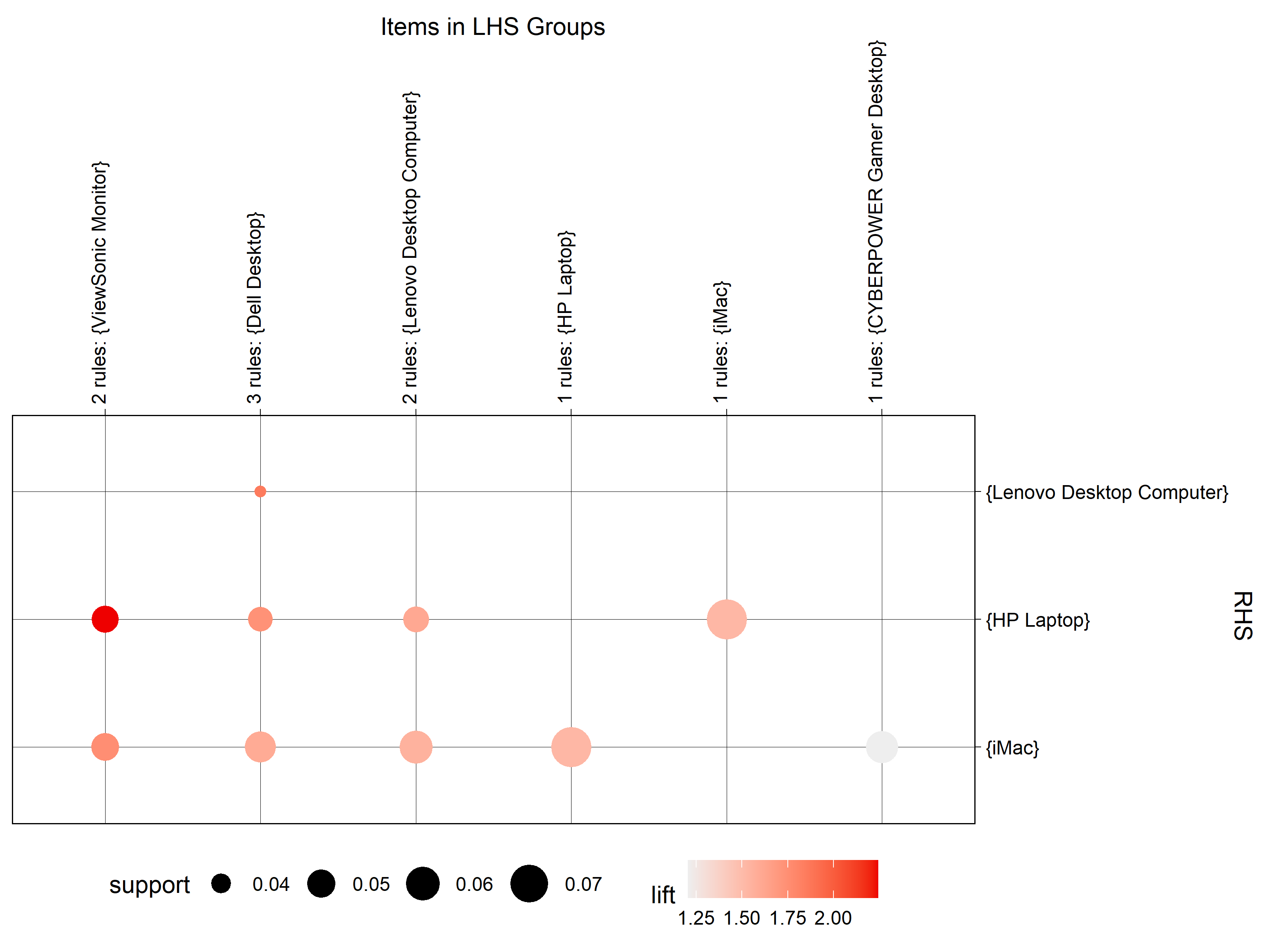
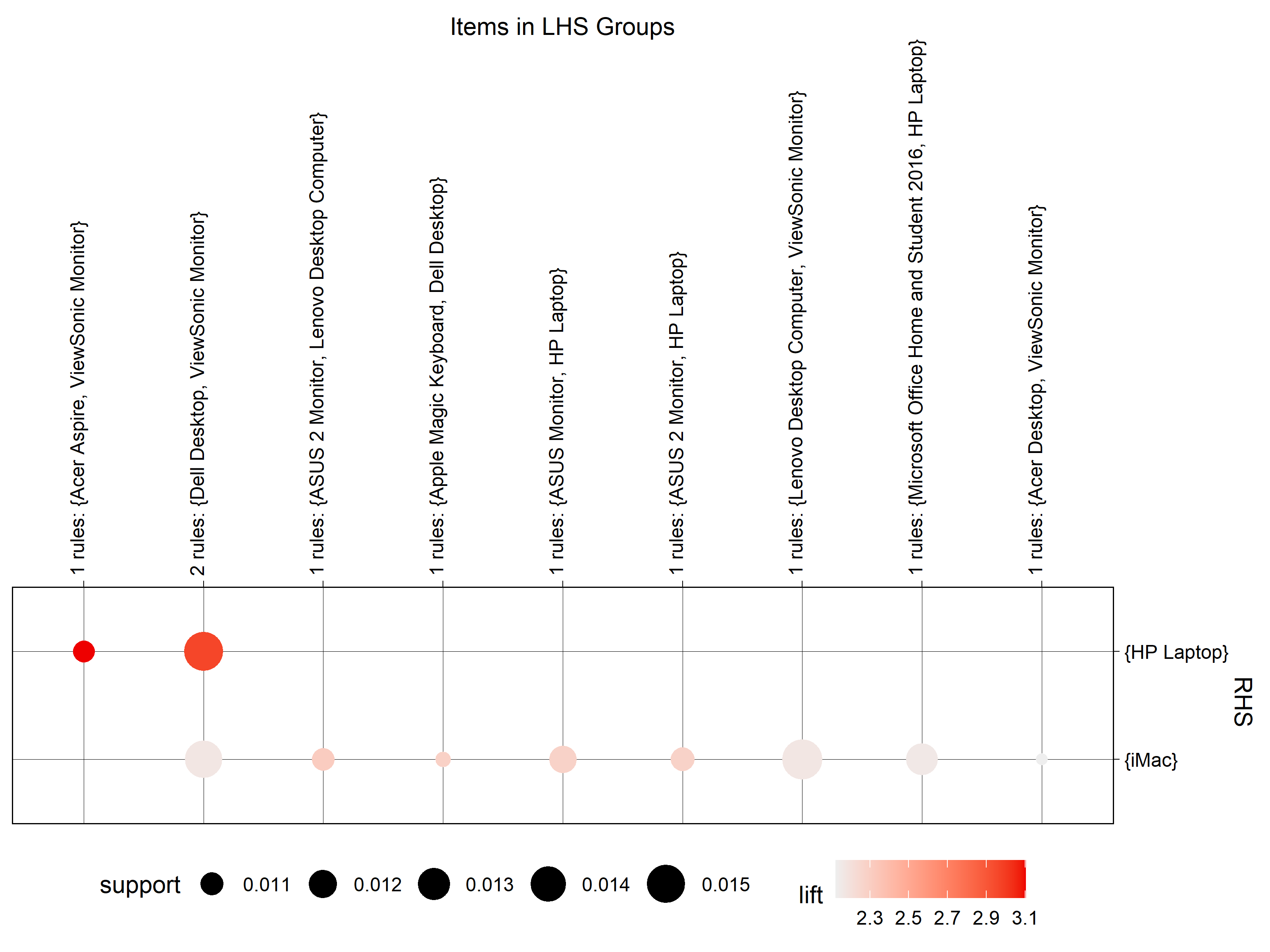


Figure 5 –Support and Lift of the Top 10 rules according to Confidence (top) and Support (bottom)

As can be seen, all these rules have got a quite positive lift increasing the amount of sales of the product on the right hand side, however these are quite unclear as mostly the rules register items of the same type. The iMac and HP Laptops are items that are so popular that cross-selling should be promoted for the clients that intend on buying one of these items.

Out of curiosity, I tried plotting the most relevant rules that did not contain iMac’s or HP Laptops and obtained the following:

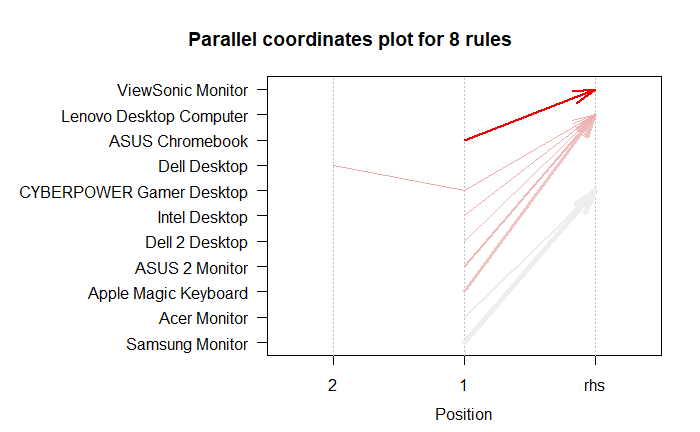


Figure 6 - Parallel coordinates for all 8 rules that didn't contain an iMac or HP Laptop.

Looking at the above graph, half of these rules seem relevant as they are mixing items from entirely different categories. It makes perfect sense that a client that purchased a Monitor also purchased a desktop or the other way around. Once again, these types of items could be cross-sold.

## Conclusions:

To answer Blackwell’s question, yes there are relevant item patterns that could be explored in order to increase or boost sales. Blackwell could benefit from selling Electronidex’s items as new rules containing items from both stores would be created, as was seen, in the electronics world many items are closely related and quite possibly dependent on each other.

In conclusion, Blackwell will benefit from acquiring Electronidex. An in depth analysis of the available items in Blackwell and Electronidex should be made, the items that were not present in this dataset could be extrapolated from the other items of the same category in order to decide which items should be cross-sold, sold in promotions be removed. Anyhow, the general rule would be to try cross-selling or create promotions on item sets containing elements from different categories. On the other hand, items of the same type of object typically should not sell so well together. Finally, certain best-selling items such as the HP Laptop or the iMac could be used to cross-sell other types of product that could in turn cross-sell other products, this could be exploited as well.