# 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:

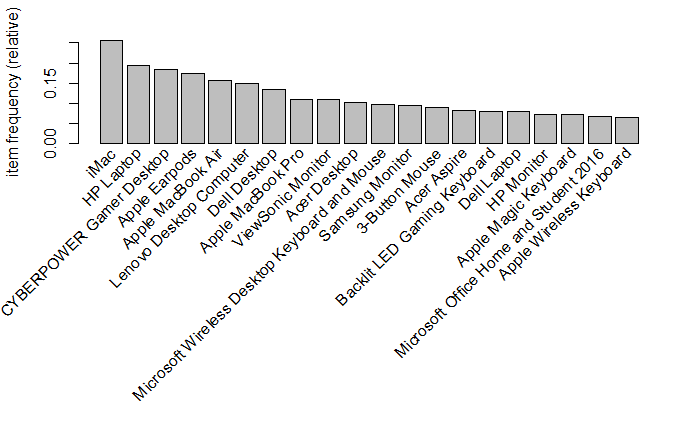


Figure 1 - Top 20 items purchased by frequency in relative terms

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:

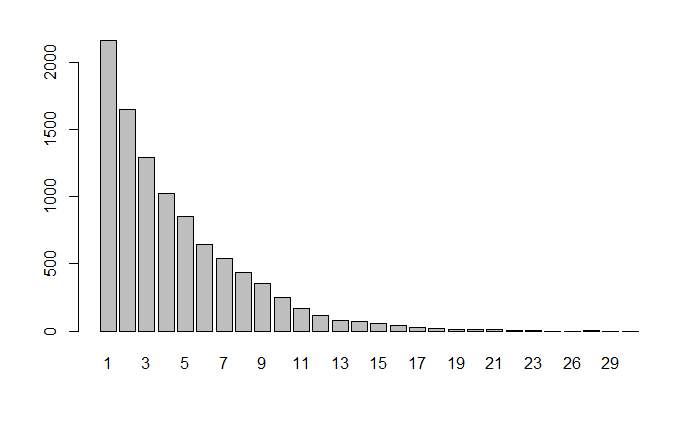


Figure 2 - Quantity of items purchased per transaction

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.

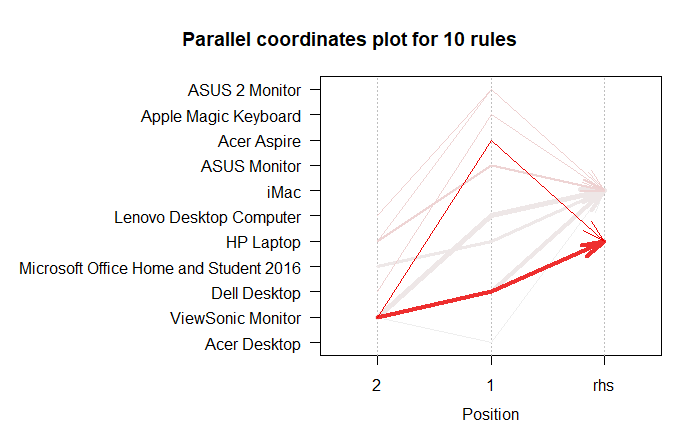


Figure 3 - Parallel coordinates for top 10 rules in term of confidence.

These types of rules are useful but not towards all types of client segments. In order to identify the most relevant rules for individual clients, it seemed more relevant to analyze rules according to support

Therefore, the top 10 rules were ordered in terms