Discover Associations Between Products

* Word Format and 1-3 page report.
* The insightful rules that you've discovered.
* Any visualizations that you've performed.
* Summary of your findings.
* Any observations you've made throughout your analysis.
* Your answers to the questions within Danielle's email.
* Any future recommendations if Blackwell was to acquire Electronidex.
* Written for a non-data audience and created for business purposes.

Main goal:

Make a customer profile and a product profile that will tell us if Blackwell should acquire Electroindex.

Conduct a **market basket analysis by** identify purchasing patterns that will provide insight into Electronidex's clientele.

Data:

**ElectronidexTransactions.csv**, is a record of one month’s (30 days’ worth) of 9835 online transactions and which items were purchased out of the 125 products Electronidex sells.

**ElectronidexItems.pdf**, is a list of the 125 products that Electronidex sells broken down into 17 product types.

Data analytics:

There may be custumer purchases and bussines purchases, so we split the data between single customers and bussines.

Transform the products into a dataframe.

TransactionNº/Item1/Item2/Item3....

* **Support:** Range/percentage(from 0 to 1) of transactions of item X
* **Confidence:** The likely of buying Y when X has been bought

Support(X U Y): Percentage of transactions of X when Y has been bought.

* **Lift**: How likely of buying Y when X has been bought considering the popularity of Y. Unlike the Confidence measurement, {Item 1} -> {Item 2} is the same as {Item 2} -> {Item 1} in the context of the lift measurement.
* >1: Likely to Y be bought with X
* <1: Unlikely to Y be bought with X
* **Conviction**(X->Y): 1-supp(Y)/1-conf(X->Y)

ALGORITHM

Support Threeshold: X% SET THE PERCENTAGE

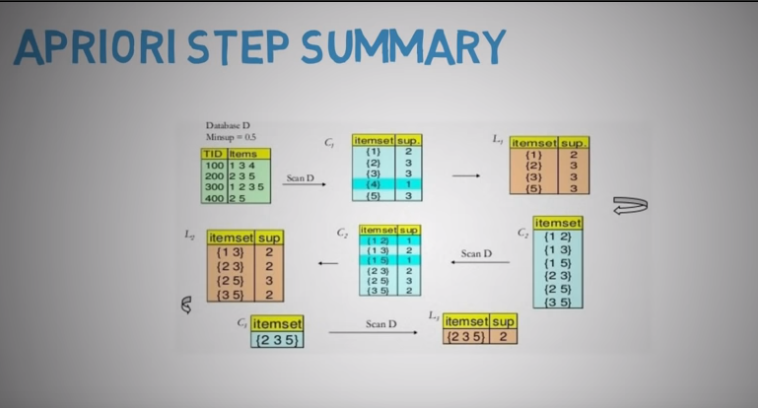
1 Set a table of items and frequency of transactions

2 set the threshold: delete the items that appear less than X%

3 Make combinations without order (ex: onions & burger=4, onions &potatoes=3....)

4 delete the ones below threshold

5 same as step 3 but with 3 items and delete the ones belwo threshold

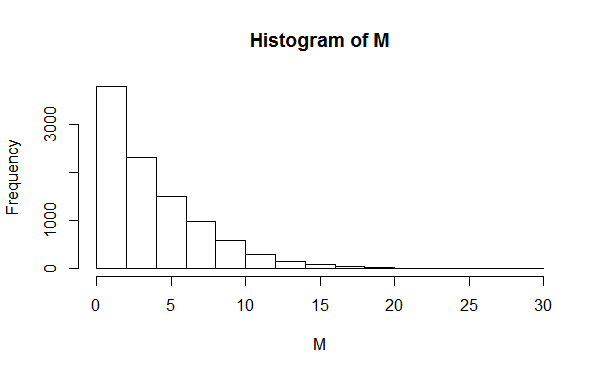


**TECHNICAL**

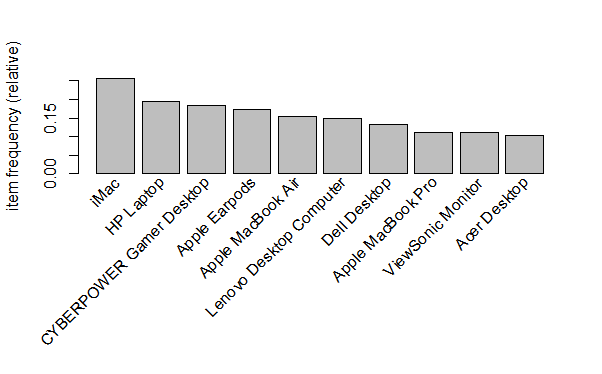
Pre-processing:

After uploading the dataset into a **sparse matrix, we see** eachrow represent a transaction and columns for each item that a customer purchased. Our inital approach could be to **split the data in two types:**

**by single costumers and bussineses.**

**Histogram of transactions.**

Most of the transactions are less than items, with 1 item being the most frequent.

**Item frequency.** Top 10 most purchased items.

**After plotting your visualizations, do you notice any patterns? Or have any observations? Take notes on your insights and observations, which might be useful to include in your formal report.**

Redundant Rules: Rules that appear repeated