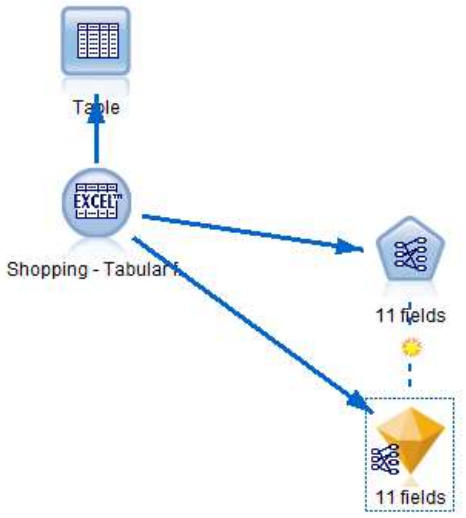
Association & Market Basket Analysis (SOLUTION)

Use SPSS Modeler and the Apriori algorithm to examine associations among transactions involving various types of books. The file (BookCLub.xlsx) contains fields that indicate whether a customer, during a single transaction, purchased a particular type of book. Thus each record represents a store visit in which at least one book was purchased.

Find useful rules considering the following thresholds:

Minimum Antecedent (LHS) Support: 10%
Minimum Rule Confidence: 80%



Sort by: Confidence % 6 of 6

Consequent	Antecedent	Support %	Confidence %	Rule Support %	Lift
CookBks	RefBks YouthBks	10.071	84.0	8.46	1.631
CookBks	YouthBks ChildBks	18.283	81.356	14.874	1.58
CookBks	YouthBks DoItYBks	12.798	81.114	10.381	1.575
CookBks	RefBks ChildBks	15.897	80.702	12.829	1.567
CookBks	RefBks DoItYBks	11.466	80.541	9.235	1.564
ChildBks	YouthBks DoItYBks CookBks	10.381	80.0	8.305	1.638

The Apriori algorithm produces a bunch of rules (see the stream attached). I considered those with improvement (aka lift) > 1.5, which means 1.5 times better than random chance, and with acceptable confidence and antecedent support (equal or above the stipulated thresholds).
In this case, for the given thresholds of LHS support and confidence, those are all 6 rules.

Antecedent		Consequent	Support %	Confidence %	Rule Support %	Lift
RefBks and YouthBks	=>	CookBks	10.071	84.00	8.46	1.631
YouthBks and ChildBks	=>	CookBks	18.283	81.36	14.87	1.58
YouthBks and DoItYBks	=>	CookBks	12.798	81.11	10.38	1.575
RefBks and ChildBks	=>	CookBks	15.897	80.70	12.83	1.567
RefBks and DoItYBks	=>	CookBks	11.466	80.54	9.24	1.564
YouthBks and DoItYBks and CookBks	=>	ChildBks	10.381	80.00	8.31	1.638

Note that I also reported Rule Support, which is a more interesting metric than Support% as defined by SPSS Modeler, which is just the support of the antecedent, or Support(LHS)

For the sake of having a better understanding of this topic, let's see how Modeler computes Lift for the first rule in the table above.

For a rule LHS = > RHS,
Lift = Prob(LHS, RHS) / Prob(LHS)*Prob(RHS), the same as
Lift = [Prob(LHS, RHS) / Prob(LHS)] / Prob(RHS), the same as

Now, Prob(LHS, RHS) / Prob(LHS) = Confidence, therefore

Lift = Confidence / Prob(RHS), the same as
Lift = Confidence / Support(RHS)

To compute Support(RHS), let us consider the first rule: **Refbks, Youthbks => Cookbks**
In this case LHS = **Refbks, Youthbks** and RHS = **Cookbks**

So Support(RHS = **Cookbks**) is computed by going to the dataset and counting how many ones there are for **Cookbks** column in all transactions.
I counted: 1662/3227 = 0.5051 = 51.50%

Therefore, Lift = Confidence / Support(RHS = Cookbks) = 84%/51.50% = 1.631

This coincides with the lift of the first rule, as calculated by Modeler.