

DSC 433/533 – Class 16 – Association Rules example for Charles case

This class exercise is based on the **Charles Book Club** Case (see separate document on the handouts page of the course website) and the Excel dataset **Charles_associations.xls** (available on the data page of the course website).

1. Find some association rules for this dataset.

- Select XLMiner > Affinity > Association Rules.
- Make sure “Charles_associations” is selected as the Worksheet, “First row contains headers” is checked, and “# rows in data” is 2000. Also, “Input data format” should be set to “Data in binary matrix format.”
- Leave the “Minimum support” at 200 and the minimum confidence at 50%.

XLMiner should find 49 rules, the first two of which are:

Rule # ↑	Conf. %	Antecedent (a)	Consequent (c)	Support(a)	Support(c)	Support(a U c)	Lift Ratio
1	100	ItalCook=>	CookBks	227	862	227	2.320186
2	62.77	ArtBks, ChildBks=>	GeogBks	325	552	204	2.274247

The first association rule is trivial (why?) but the second may be useful: “if a customer has purchased art books and children’s book, then the customer also purchases geography books.” Three measures are traditionally used to describe how good an association rule is:

- *support* (labeled “Support(a U c)” in the table)
- *confidence* (labeled “Conf. %” in the table)
- *lift* (labeled “Lift Ratio” in the table)

Using the fact that 325 customers bought art books and children’s books, 552 customers bought geography books, 204 customers bought art books, children’s books, and geography books, and there were 2000 customers in total, write out calculations to show that the second rule has confidence=62.77% and lift=2.274.

$Confidence = support(a \cup c) / support(a) = 204 / 325 = 62.77\%$,

$and\ lift = confidence / [support(c) / 2000] = 0.6277 / [552 / 2000] = 0.6277 / 0.276 = 2.274$,

$or\ lift = support(a \cup c) / [support(a) * support(c) / 2000] = 204 / (325 * 552 / 2000) = 204 / 89.7 = 2.274$.

2. Change the minimum support and confidence thresholds to find additional decision rules. In particular:

- Change the “Minimum support” to 100 transactions and the minimum confidence to 90%.

10 rules:

Rule # ↑	Conf. %	Antecedent (a)	Consequent (c)	Support(a)	Support(c)	Support(a U c)	Lift Ratio
1	100	ItalCook, YouthBks=>	CookBks	118	862	118	2.320186
2	100	GeogBks, ItalCook=>	CookBks	128	862	128	2.320186
3	100	DoltYBks, ItalCook=>	CookBks	117	862	117	2.320186
4	100	ArtBks, ItalCook=>	CookBks	113	862	113	2.320186
5	100	ItalCook=>	CookBks	227	862	227	2.320186
6	100	ChildBks, ItalCook=>	CookBks	170	862	170	2.320186
7	100	ChildBks, GeogBks, ItalCook=>	CookBks	105	862	105	2.320186
8	91.38	DoltYBks, RefBks, YouthBks=>	ChildBks	116	846	106	2.160267
9	90.35	GeogBks, RefBks, YouthBks=>	ChildBks	114	846	103	2.135955
10	90.27	ChildBks, DoltYBks, GeogBks, YouthBks=>	CookBks	113	862	102	2.094327

- Change the “Minimum support” to 60 transactions and the minimum confidence to 20%.

1343 rules, including 3 with Florence as the result/consequent:

Rule #	Conf. %	Antecedent (a)	Consequent (c) ↑	Support(a)	Support(c)	Support(a U c)	Lift Ratio
953	23.53	ArtBks, GeogBks=>	Florence	255	217	60	2.168609
1219	20.62	ArtBks, ChildBks=>	Florence	325	217	67	1.900035
1258	20.12	ArtBks=>	Florence	482	217	97	1.854791