In-Class Exercise: Computing Confidence, Support, and Lift

Here are the baskets from eight shoppers:

Basket	Items		
1	Coke, Pop-Tarts, Donuts		
2	Cheerios, Coke, Donuts, Napkins		
3	Waffles, Cheerios, Coke, Napkins		
4	Bread, Milk, Coke, Napkins		
5	Coffee, Bread, Waffles		
6	Coke, Bread, Pop-Tarts		
7	Milk, Waffles, Pop-Tarts		
8	Coke, Pop-Tarts, Donuts, Napkins		

Compute the support, confidence, and lift for the following rules:

Rule		Support	Confidence	Lift	
1	{Coke, Pop-Tarts} \rightarrow {Donuts}	2/8 = 0.25	2/3 = 0.67	$\frac{0.25}{0.25} = 1.78$	
2	{Coke} \rightarrow {Pop-Tarts, Donuts}	2/8 = 0.25	2/6 = 0.33	0.375 * 0.375 0.25	
2	(Cake) N (Denute Neuline)	2/8 - 0.25	2/6 - 0.22	$\frac{1}{0.75 * 0.25} = 1.33$	
3	{Coke} \rightarrow {Donuts, Napkins}	2/8 = 0.25	2/6 = 0.33	$\frac{0.23}{0.75 * 0.25} = 1.33$	
4	{Coffee} \rightarrow {Bread, Waffles}	1/8 = 0.125	1/1 = 1.00	$\frac{0.125}{0.125 * 0.125} = 8.0$	
5	${Coke} \rightarrow {Donuts}$	3/8 = 0.375	3/6 = 0.50	$\frac{0.375}{0.75 * 0.375} = 1.33$	

1) Which rule has the strongest association? How do you know?

Rule 4: {Coffee} → {Bread, Waffles}

It has the highest lift (8.0) and the highest confidence (1.00).

2) Consider a customer who is walking through the store with only a bottle of coke in their shopping cart. You then see them put pop-tarts in their cart. Do you become more or less sure than you were before that they will buy donuts? Explain.

You become more sure because the lift value goes up. {Coke} \rightarrow {Donuts} has a lift of 1.33 {Coke, Pop-Tarts} \rightarrow {Donuts} has a lift of 1.78

Computing lift based on aggregate purchase numbers

1) Consider two products, the Squishee and the Peanut Butter Bowl. Here's a profile of 18,500 customers:

Squishee

		No	Yes	
Peanut	No	10000	2000	
Butter Bowl	Yes	1500	5000	
				18500

What is the lift for the rule {Peanut Butter Bowl} \rightarrow {Squishee}? (Are people who bought a Peanut Butter Bowl more likely than chance to buy a Squishee too?)

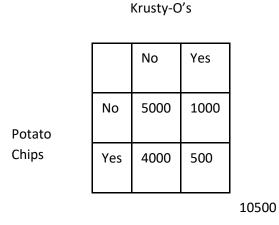
Support(Peanut Butter Bowl, Squishee) = 5000/18500 = 0.270 Support(Peanut Butter Bowl) = 6500/18500 = 0.351 Support(Squishee) = 7000/18500 = 0.378

$$Lift = \frac{s(Peanut Butter Bowl, Squishee)}{s(Peanut Butter Bowl) * s(Squishee)} = \frac{0.27}{0.351 * 0.378} = 2.03$$

So these products are strongly associated – they appear in the same basket more often than what you'd expect by chance (i.e., Lift > 1). If a customer buys a Peanut Butter Bowl, they are likely to also buy a Squishee.

The numbers may be off very slightly due to rounding.

2) Consider two products, Potato Chips and Krusty-O's. Here's a profile of 10,500 customers:



What is the lift for the rule {Potato Chips} \rightarrow {Krusty-O's}? (Are people who bought Potato Chips more likely than chance to buy Krusty-O's too?)

Support(Potato Chips, Krusty-O's) = 500/10500 = 0.048 Support(Potato Chips) = 4500/10500 = 0.429 Support(Krusty-O's) = 1500/10500 = 0.143

 $Lift = \frac{s(Potato\ Chips, KrustyOs)}{s(Potato\ Chips) * s(KrustyOs)} = \frac{0.048}{0.429 * 0.143} = 0.782$

So these products have a negative association – they appear in the same basket less often than what you'd expect by chance (i.e., Lift < 1). If a customer buys Potato Chips, they are unlikely to also buy Krusty-O's.

The numbers may be off very slightly due to rounding.