**In-Class Exercise: Understanding Association Mining**

***Part 1: Reading Association Mining Output***

The following is the output from an association mining analysis using SAS Enterprise Miner. The data set captures the parts of the site used by a radio station’s visitors over a two-month period.

Those services are:

ARCHIVE Archives of past news stories
WEBSITE General information on the site
MUSICSTREAM Internet streaming of the station (audio-only)
SIMULCAST Video streaming of live music performances
PODCAST Downloading podcasts of individual programs

The key elements from the Rules Table are provided below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Confidence** | **Support** | **Lift** | **Transaction Count** | **Rule** |
| 45.18 | 7.05 | 1.67 | 111878 | ARCHIVE ==> WEBSITE |
| 12.26 | 7.05 | 1.67 | 111878 | WEBSITE ==> ARCHIVE |
| 96.31 | 5.35 | 0.79 | 84789 | MUSICSTREAM ==> WEBSITE |
| 43.27 | 4.1 | 0.75 | 64964 | SIMULCAST ==> WEBSITE |
| 26.64 | 15.32 | 0.61 | 243066 | WEBSITE ==> PODCAST |
| 35.08 | 15.32 | 0.61 | 243066 | PODCAST ==> WEBSITE |

Now answer the following questions:

1. Which rule(s) have the highest confidence?
2. Which rule(s) have the highest support?
3. Which rule(s) have the highest lift?
4. What are the two rule “pairs” in the list above?
5. What other service “goes the most” with visiting the website for general information (WEBSITE)? In other words, what other service are WEBSITE visitors most likely to seek out? What statistic did you use to figure this out?
6. What other service seems to “go the least” with visiting the website for general information (WEBSITE)? In other words, what other service are WEBSITE visitors least likely to seek out? What statistic did you use to figure this out?
7. The rule MUSICSTREAM ==> WEBSITE has poor lift (i.e., less than 1), but the rule has the highest confidence. Explain how this is possible.

***Part 2: Computing Support, Confidence, 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} 🡪{Donuts} |  |  |  |
| 2 | {Coke} 🡪 {Pop-Tarts, Donuts} |  |  |  |
| 3 | {Coke} 🡪 {Donuts, Napkins}  |  |  |  |
| 4 | {Coffee} 🡪 {Bread, Waffles} |  |  |  |
| 5 | {Coke} 🡪 {Donuts} |  |  |  |

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

1. 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.

***Part 3: 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 |  |
| Peanut Butter Bowl |  | No | Yes |  |
| No | 10000 | 2000 |  |
| Yes | 1500 | 5000 |  |
|  |  |  |  | 18500 |

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

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

|  |  |  |
| --- | --- | --- |
|  | Krusty-O’s |  |
| Potato Chips |  | No | Yes |  |
| No | 5000 | 1000 |  |
| Yes | 4000 | 500 |  |
|  |  |  |  | 10500 |

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