In-Class Exercise: 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?
   MUSICSTREAM ==> WEBSITE

2) Which rule(s) have the highest support?
   WEBSITE ==> PODCAST and PODCAST ==> WEBSITE

3) Which rule(s) have the highest lift?
   ARCHIVE ==> WEBSITE and WEBSITE ==> ARCHIVE

4) What are the two rule "pairs" in the list above?
   ARCHIVE ==> WEBSITE/WEBSITE ==> ARCHIVE and
   WEBSITE ==> PODCAST/PODCAST ==> WEBSITE

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?
   ARCHIVE – LIFT is greater than 1. This implies that this isn't just random chance – people are actively seeking out the WEBSITE if they've used the ARCHIVE.

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?
    **PODCAST – LIFT is less than 1. This also implies that this isn't just random chance – but
    this time, people who visit the web site are particularly unlikely to also download a
    podcast.**

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.
    **It could be that many people use both MUSICSTREAM and WEBSITE so it appears in
    visitors' set of services a lot. However, there can still be a negative effect of one on the
    other. For example, I use the website a lot, and I use music streaming a lot, but I'm still
    less likely to do one if I've done the other – possibly they are substitutes.**