**Assignment #9 – SAS #4: Association Rules**

 **Guidelines**

* Follow the steps below, using the in-class exercise as a guide.
* You must submit your answers electronically in a single Word document. Fill in the answer sheet at the end of this document.
* You must include your name at the top of the document.
* Your answers should be emailed, as an attachment, to your instructor with the subject:
 **MIS2502: Association Rules**
* The email must be sent by the start of class the day the assignment is due.

***If you do not follow these instructions, your assignment will be counted late.***

A store is interested in determining the associations between items purchased from the Health and Beauty Aids department and the Stationery Department. The store chose to conduct a market basket analysis of specific items purchased from these two departments. The **TRANSACTIONS** data set contains information about over 400,000 transactions made over the past three months. The following products are represented in the data set:

* bar soap
* bows
* candy bars
* deodorant
* greeting cards
* magazines
* markers
* pain relievers
* pencils
* pens
* perfume
* photo processing
* prescription medications
* shampoo
* toothbrushes
* toothpaste
* wrapping paper

There are four variables in the data set:

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Model Role** | **Measurement Level** | **Description** |
| **STORE** | Rejected | Nominal | Identification number of the store |
| **TRANSACTION** | ID | Nominal | Transaction identification number |
| **PRODUCT** | Target | Nominal | Product purchased |
| **QUANTITY** | Rejected | Interval | Quantity of this product purchased |

* 1. Create a new diagram. Name the diagram **Transactions**.
	2. Create a new data source using the data set **AAEM.TRANSACTIONS**.

	At step 6 of the data source wizard, assign the variables **STORE** and **QUANTITY** the model role **Rejected**. These variables will not be used in this analysis.

	Assign the ID model role to the variable **TRANSACTION** and the Target model role to the variable **PRODUCT**. Make sure that both TRANSACTION, PRODUCT, and STORE are set to Nominal for Measurement Level.
	 **(BE CAREFUL HERE! MAKE SURE THE SETTINGS MATCH THE TABLE ABOVE!)**

	On step 9, make sure the role is set to **Transaction**.
	3. Add the node for the **TRANSACTIONS** data set and an Association node to the diagram. Connect the data set node to the Association node.
	4. In the Properties tab for the Association node, change the setting for Export Rule by ID to **Yes**.
	5. Leave the remaining default settings for the Association node and run the node.
	6. Examine the results of the association analysis
	7. What is the highest lift value for the set of rules?

**Waypoint:** The support for these rules is 2.18.

**ANSWER:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (use the answer sheet)

* 1. Which rule pair has that highest lift value?

	**ANSWER:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (use the answer sheet)

**Include a screen shot showing where you found your answer (use the answer sheet)**

* 1. Your manager comes to you and insists that you put the toothbrushes next to the toothpaste because they “go together” (in other words, they are likely to be purchased together). Is that supported by the data? Briefly explain your answer.

	ANSWER: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (use the answer sheet)
	2. List two products that are positively associated with toothpaste. Include the lift statistic used to reach this conclusion.

	**ANSWER (product 1): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** (use the answer sheet)
	**ANSWER (product 2): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** (use the answer sheet)
	(make sure you include the lift statistic in your answers)
	3. Look at rules 32 and 33 (Magazine🡪Toothpaste and Toothpaste🡪Magazine)? Describe the association between magazines and toothpaste, using the lift statistic to support your answer.
	 **ANSWER: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** (use the answer sheet)
	(make sure you include the lift statistic in your answer)

You won’t be using SAS for these last two:

* 1. Consider the following set of customer service visits for an auto repair shop:

|  |  |
| --- | --- |
| **Visit** | **Services Performed** |
| 1 | Oil Change, Tire Rotation, Brake Service |
| 2 | Oil Change, Tire Rotation |
| 3 | Filter Replacement, Tire Rotation |
| 4 | Brake Service, Oil Change |
| 5 | Filter Replacement, Oil Change, Brake Service |

Compute support, confidence, and lift for the following rules (use the answer sheet):

|  |  |  |  |
| --- | --- | --- | --- |
| **Rule** | **Support** | **Confidence** | **Lift** |
| 12a | {Oil Change} 🡪 {Brake Service} |  |  |  |
| 12b | {Brake Service} 🡪 {Tire Rotation, Oil Change} |  |  |  |
| 12c | {Filter Replacement} 🡪 {Brake Service}  |  |  |  |

* 1. SchuffMart! has started carrying two new products: QuirkyJerky, a soy-based non-meat beef jerky, and GreenBull, an energy drink made entirely of kelp. After six months they created the following analysis of sales from 25,800 total customers:

|  |  |  |
| --- | --- | --- |
|  | **Bought GreenBull** |  |
| **Bought QuirkyJerky** |  | **No** | **Yes** |  |
| **No** | 7500 | 8500 |  |
| **Yes** | 5300 | 4500 |  |

Are people who buy QuirkyJerky inclined to buy GreenBull at a greater rate than what would occur by chance? Support your answer by providing the lift value for the rule:
{ QuirkyJerky } => { GreenBull }.

**ANSWER:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (use the answer sheet)
(make sure you include the lift statistic in your answer)

Answer Sheet for Assignment 9

Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

 *Note that the item numbers correspond to the assignment instructions.*

|  |  |
| --- | --- |
| Item Number | Answer |
| 7 |  |
| 8 (rule 1) |  |
| 8 (rule 2) |  |
| 8 (screen shot) |  |
| 9 |  |
| 10 (product 1) |  |
| 10 (product 2) |  |
| 11 |  |
| 12a (support) |  |
| 12a (confidence) |  |
| 12a (lift) |  |
| 12b (support) |  |
| 12b (confidence) |  |
| 12b (lift) |  |
| 12c (support) |  |
| 12c (confidence) |  |
| 12c (lift) |  |
| 13 |  |