**Assignment #9 Association Rules**

|  |
| --- |
| **Submission Instructions****Deadline:** **December 7, 2018, midnight**.* Submit the following **three** files through Blackboard:
1. The completed, working **R script** that produced the analysis.
2. The **output file** “ARulesOutput.csv.”
3. The completed **answer sheet** provided on the last page.
* If you do not follow the instructions, your assignment will be counted late.
	+ Late Assignment policy: Same as before.

**Evaluation**Your submission will be graded based on the correctness of the completed answer sheet, with other files as supporting documents. |

**Before you start**

For this assignment, you’ll be working with the **Groceries.csv** file and the **aRules.r** script. This file reflects 43,366 items purchased across 9,835 transactions at a grocery store. Each row represents a transaction/product pair. The first value is the transaction ID and the second value is the product name (such as citrus fruit, yogurt, coffee, etc.). The manager of the store wants to better understand his customer’s buying habits so that he can modify the store layout and run more effective promotions.

**Guideline**

Leave the default settings in the **aRules.r** script in place, but make sure you change the **input file name** before you run the analysis.

**Part 1. Based on the output from R, answer the following questions (by completing the answersheet at the end of this document):**

1. Open the **ARulesOutput.csv** file generated after running aRules.r. Before sorting the results, look at rules 538 and 580, where the corresponding rules are { other vegetables, whole milk} => { whipped/sour cream} and { other vegetables, yogurt} => { root vegetables}.

Given their almost identical lift values, what statistic would you use to determine which one had the stronger association?

1. What is the highest lift value in the list of rules? Which association rule has this highest lift value? *(Hint: You need to look at the ARulesOutput.csv file. Sort the table in descending order by lift.)*
2. Your store manager comes to you and suggests that they should place the soda and citrus fruit together because, as beverages, they “go together.” Is that supported by the data? Cite the statistic(s) you used and the value(s) you used to reach this conclusion.

*(Hint: look for rules { citrus fruit} => { soda} and { soda} => { citrus fruit}.)*

1. You decide to start a promotion that will help cross-sell yogurt. Which product(s) have the highest predictive power in determining who will purchase yogurt? Cite the statistic(s) you used, and the value(s), to reach this conclusion.

**Part 2. Answer the following questions based on calculation. You won’t be using R for the following problems:**

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, Filter Replacement |
| 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** |
| a | {Oil Change} 🡪 {Brake Service} |  |  |  |
| b | {Brake Service} 🡪 {Tire Rotation, Oil Change} |  |  |  |

1. The store 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** | 7000 | 8500 |  |
| **Yes** | 5300 | 5000 |  |

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 Sheet for Assignment: Association Rules

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

 *Fill in the answersheet below with the answers to the questions on pages 1-2 of the assignment:*

|  |  |
| --- | --- |
| Question | Answer |
| **Part 1. Based on the output from R, answer the following questions** |
| 1 |  |
| 2 |  |
| 3 |  |
| 4 |  |
| **Part 2. Answer the following questions based on calculation. You won’t be using R for the following problems:** |
| 5a | Support:Confidence:Lift: |
| 5b | Support:Confidence:Lift: |
| 6 | Yes or no:Lift: |