**Assignment #4 Association Rules**

|  |
| --- |
| **Submission Instructions**   * Submit the followingfiles through Blackboard:  1. The **output file** “association\_rules.xlsx.” 2. The completed **answer sheet** provided on the last page.  * If you do not follow the instructions, your assignment will be counted late. |

**Before you start**

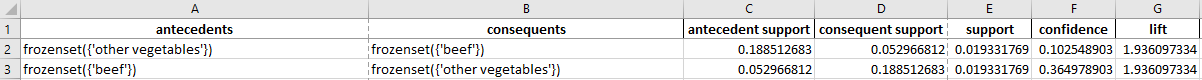
For this assignment, you’ll be working with the **Groceries2.csv** file and the **Lab4.py** script. This file reflects 18,996 items purchased across 8,948 transactions at a grocery store. Each row represents a transaction/product pair. The manager of the store wants to better understand his customer’s buying habits so that he can modify the store layout and run more effective promotions.

**Guideline**

Leave the default settings in the **Lab4.py** script in place. Only change the **dataset** before you run the analysis.

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

1. Open the association\_rules.xlsx file generated after running the code. Before sorting the results, look at rules #1 and #2:



Which one had the stronger association? Why?

1. Which association rule has the highest lift value?
2. Your store manager comes to you and suggests that they should place the “soda” and “fruit/vegetable juice” 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.
3. 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.
4. From the results of association rule mining, do you have any other interesting findings? List three of them and explain how the manager can use these findings to improve the business.

**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} | **3/5** | **3/4** |  |
| 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 | Total: 25800 |

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 }.C

Answer Sheet for Assignment: Association Rules

*Fill in the answer sheet 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 | Rule 1: 'Beef' -> 'Other Vegetables' rule (Lift: 1.936097334)  Rule 2: 'Other Vegetables' -> ‘Beef’ rule (Lift: 1.936097334)  The same lift value, 1.936097334, for both rules indicates that the association's strength is the same in both directions. In this situation, the link between "beef" and "other vegetables" is strong regardless of the association's direction because both rules have the same lift. |
| 2 | ‘Other vegetables', 'tropical fruit’ -> ‘root vegetables’ |
| 3 | According to the confidence value (0.103623649), 10.36% of transactions that include "soda" also include "fruit/vegetable juice." The support, confidence, and lift values are also relatively low. Thus we can say that there is a weak association between "soda" and "fruit/vegetable juice." Due to this, grouping them together based only on their beverage category may not be the best choice even though they are not closely related. |
| 4 | Based on the lift values, we can say that the following products have the highest predictive power in determining who will purchase yogurt:   * Curd (lift: 3.38407563) * Root vegetables (lift: 2.846249177) * Curd and whole milk (Lift 2.8231768239) * Tropical fruit (lift: 2.70692745)   These products have a strong positive association with yogurt, indicating that customers who purchase these items are more likely to also purchase yogurt. |
| 5 | Association between "other vegetables" and "tropical fruit" with "root vegetables"  Association between "yogurt" and "whole milk" with "curd"  Association between "beef" and "root vegetables"  The highest lift values are found for these three connections. The manager can carefully organize merchandise, develop alluring promos, and give customers a smooth shopping experience by utilizing these extra discoveries. Understanding these relationships enables the manager to customize their product offers and marketing methods, ultimately enhancing business performance. |
| **Part 2. Answer the following questions based on calculation. You won’t be using R for the following problems:** | |
| 6 | {Oil Change} 🡪 {Brake Service}  Support- 3/5=0.6  Lift- 0.6 / (4/5 \* 3/5) = 1.25  Confidence- 3/4= 0.75  {Brake Service} 🡪 {Tire Rotation, Oil Change}  Support 1 / 5 = 0.2  Lift 0.2 / (0.6 \* 0.4)) = 0.8333  Confidence 1 / 3 = 0.33 |
| 7 | Support ({QuirkyJerky} => {GreenBull}) = 5000/25800= 0.193  Support (QuirkyJerky) = 10300/25800= 0.399  Support (GreenBull) = 13500 / 25800= 0.523  Lift= 0.193/ (0.399\*0.523) = 0.9248  The lift value in this instance is less than 1, indicating that customers who purchase QuirkyJerky are less likely to purchase GreenBull than would happen by chance. |