MIS2502: Data Analytics for Management
Assignment: Association Rule Mining Using R

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. 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.

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

Answer the following questions (complete the worksheet at the end of this document):

1) What is the highest lift value in the list of rules?
2) Which rule has this value? (write the number of the rule and the rule itself)
3) 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 you used, and its value, you used to reach this conclusion.
4) You decide to start a promotion that will help cross-sell butter. Which product(s) have the highest predictive power in determining who will purchase butter? Cite the statistic you used, and its value, to reach this conclusion.
5) Look at rules 538 and 580. Given their almost identical lift values, what statistic would you use to determine which one had the stronger association?

You won't be using R for these last two problems:

6) 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
6a	{Oil Change} → {Brake Service}			
6b	{Brake Service} → {Tire Rotation, Oil Change}			
6c	{Filter Replacement} → {Brake Service}			

7) 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	
		No	**Yes**
Bought QuirkyJerky	**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 }.

What to submit:

Send a single email to your instructor with the following attachments:

- The completed, working R script that produced the analysis.
- The output file "ARulesOutput.txt."
- The completed worksheet provided on the next page.

Answer Sheet for Assignment: Association Mining Using R

Name _____

Fill in the worksheet below with the answers to the questions on page 2 of the assignment:

Question	Answer
1	
2	
3	
4	
5	
6a	Confidence: Support: Lift:
6b	Confidence: Support: Lift:
6c	Confidence: Support: Lift:
7	