**Part 2 - Run an exercise on the credit approval dataset you used for the last week exercise and write a report on your findings in your own words. The report needs to cover the exercise key points below in order.  Save all commands in an R script.**

1. Introduction - What do you expect the Apriori method to accomplish for credit approval data?

2. Data Pre-processing

Load the Credit Approval data in RStudio.  You may use the tools menu or you may run read.csv command.

What data pre-processing does the Apriori method require for Credit Approval data?Include the commands you ran, the output, and the output interpretation in the report.  For each command, explain its purpose.

3. Run the method with default arguments and store the generated rules in a variable called rules.

1. Include the command, the output, and the output interpretation in the report.  Discuss the number of returned rules and the default arguments.
2. Run the inspect command to  display the first 10 rules and interpret the output, including the returned rules and metrics.  Include the command, the output, and interpretation in the report

4. Run the method with 2 different combinations of confidence, support, and minimum length values.

1. For each run
   * Specify the input parameters you used
   * Include the command and the command output.
   * Discuss how many rules and how many items were returned
   * Run the inspect command to preview the first 10 rules.  What is the strongest rule, and why? Include the command, the output, and output interpretation in the report.
2. How does changing confidence, support, and minimum length values affect the returned rules?
3. What are the differences between support, confidence, and lift metrics for identifying the strongest rules?

5. Generate the rules that have only the class=’+’ or class=’-’ on the right hand side.  Store the rules in the variable rules. (Hint - See the generating rules for the specified itemsets section in the Word Document. You may need to adjust the confidence and support values.)

1. Include the command, the output, and output interpretation in the report.
2. Run the inspect command to preview the first 10 rules.  Include the command and output in the report.
3. What do the returned rules suggest about the credit approval decision? What are the strongest rules?

6. Prune the returned rules

1. Why do we prune the returned rules?
2. Run the following commands on a variable that stores the rules class=’+’ or class=’-’ on the right hand side to find the redundant rules and to eliminate them.  Discuss the output of which(redundant) command.  Include the output and discussion in the report.

rules.sorted <- sort(rules, by="lift")

inspect(rules.sorted)

subset.matrix <- is.subset(rules.sorted, rules.sorted)

subset.matrix[lower.tri(subset.matrix, diag=T)] <- NA

redundant <- colSums(subset.matrix, na.rm=T) >= 1

which(redundant)

1. Run the following commands to remove the redundant rules and to display the remaining rules.  Which rules remain? Include the commands, the output, and output interpretation in the report.

rules.pruned <- rules.sorted[!redundant]

inspect(rules.pruned)

7.  Rules visualization -  Choose any visualization method discussed in the tutorial to visualize the pruned rules in step 6.  Explain how the plot represents the rules and the metrics for ranking rules.  How do we use the plot to identify the strongest rules? Include the commands, the plot, and plot discussion in the report.

8.  Summary

1. Why do we consider more than one metric to identify the strongest rules?
2. Which part of this exercise did you find the most challenging, and why? What approach did you take to resolve the challenge?

**Optional** - An example on page 13-14 shows how to display additional rule interest measures, including leverage, conviction, and coverage. What additional information do those metrics provide about the rules, and why would we consider them? You may find the following article useful.  [Additional Metrics for Association Rules](http://www.cse.msu.edu/~ptan/papers/IS.pdf)

(Although additional metrics exploration is optional and will not be part of the assignment grade, it will enhance an understanding of the topic.)

**Exercise Deliverable**

Submit the following files in the Week 3 Exercise Assignment folder.

* The report addressing the key points above in order and in your own words
* An R script with commands your ran and brief comments on the commands purpose