

DSCI 300 - Assignment 6

Instructions

In the space below, answer each of the following questions. Please submit your final assignment as a .pdf file created in Markdown. Please delete this section of your document prior to submission. You should change the author and date above before you begin. Feel free to give the assignment your own title.

Your submission should be in 3 parts:

- Part A includes a written report (with graphics and figures, if appropriate) that would be suitable for submission to the client. This report should not include any R code or question numbers. It should be well-formatted and address all of the questions asked below.
- Part B (Appendix A) includes all of the answers to each of the questions below, broken down by question number. This portion should include all of the R code needed to complete part A. Essentially, part B is the rough work required to make part A (often, we include this sort of an analysis as an appendix for the client).
- Part C (Appendix B) includes the codebook for the dataset.

Association Rules of Grocery Store Transactions

A grocery store introducing items from Italy is interested in analyzing buying trends of these new “international” items, namely prosciutto, Peroni, risotto, and gelato. The files GroceryStoreList and GroceryStoreStacked provide data on a collection of transactions in item-list format.

Managerial Report

- a. Rather than a codebook in Appendix B, you are to clean the data and import it into R in whatever way you wish. This could be as painstaking as writing each individual transaction out or as simple as finding a clever way to read the data. Whatever you decide, replace the codebook with a set of instructions for someone to import the data into R the same way that you did. Note: there are some very easy and unconventional ways to read this data into R.
- b. Create a frequency distribution for the new Italian items.
- c. Use a minimum support of 25% of the transactions (250 of the 1,000 total transactions) and a minimum confidence of 50% to generate a list of association rules. How many rules satisfy this criterion? Why may the grocery store want to change the minimum support required for their analysis? What is the risk of changing the minimum support required?
- d. Using the list of rules from part (b), consider the rule with the largest lift ratio that also involves an Italian item. Interpret what this rule is saying about the relationship between the antecedent item set and consequent item set.
- e. Interpret the confidence of the rule with the largest lift ratio that also involves an Italian item.
- f. Interpret the lift ratio of the rule with the largest lift ratio that also involves an Italian item.
- g. What insight can the grocery store obtain about its purchasers of the Italian fare?
- h. The store wants to move online and wants to create a recommender system for these Italian items. For each Italian item, determine the top 5 additional items that the store should recommend to the customer.