Data Mining Homework 2

Shopping data from 10,000 grocery customers:

B=bread, J=jelly, P=peanut butter

Assume, for example, that "B 500" means 500 people bought bread

possibly with the other items, possibly not. Likewise P & J

incudes those buyers of peanut butter and jelly who also bought

bread as well as those buyers of peanut butter and jelly who

didn't (etc.)

B 500

J 400

P 300

B & J 180

P & J 140

B & P 120

(1) Assuming 80 people bought all 3, find the support, confidence,

and lift for these rules. Report the results

in a nice table.

P&J => B

B => P&J

P => B

P => not B

(2) What are the largest and smallest possible values for the number

of people who bought all three products?

FOr the smallest possible value, what would be the the support,

confidence, and lift for the rule P => J ? How would this change

if we changed the number of people who bought all three products?

(3) (Optional) Make a table, or better yet a graph of Y=lift

for the rule which states that customers who buy both bread and

peanut butter will also buy jelly. Do this for all possible values

between the two extremes from question (2) above. For a graph

the horizontal axis is the number customers who bought all 3

items and for a table, that number labels the rows of the table.

Can the graph be described in a simple sentence?

(4) Work through the association analysis exercise (Exercise 2

page 8-82) concerning sales of health and beauty aids and stationery

at a store. The transactions dataset is on our home page under

data. Answers to book exercises are in the book so I am mainly

looking for a nice writeup of the results.