Shopping data exercises

To answer these questions, you will need a data set (available on the module notes site), which is the Excel spreadsheet named with your student username. You will use this same file for part of the assignment, so store it somewhere safe.

Background

You work in data analytics for a supermarket. Your employer has provided you with a data set of customer purchases. Some notes on these data:

- You will see that you have data on purchases from ten customers in ten worksheets.
- The first column indicates week number.
- The items purchased are indicated in columns.
- The first row is the name of the item for sale.
- The second row gives the price of the item.
- After this, rows use '1' to indicate a purchase this week and '0' to indicate no purchase this week.
- The columns are ordered alphabetically.

Questions

1. There are three types of apple for sale in the supermarket. Find these for customer 1 (hint: use the 'find' function to search for the word "apple"). Use the data from customer 1 to generate a matrix of transition probabilities for the three types of apple. What are the probabilities of customer 1 purchasing each type of apple in the week following the data collection?

2. There are two types of rice pudding for sale. Count how many times each type of rice pudding is bought by each customer, and use these counts to estimate the market share of each type of rice pudding.

   One method could be: for customer 1, find the two columns that represent rice pudding and hide all the other columns; now in column BL make counts of each type of rice pudding; then copy column BL to each of the other sheets in the spreadsheet (they should update to give a count of that customer's rice pudding data). Now you can sum all the purchases and calculate what percentage were of each type of rice pudding.

3. Your employer wishes to understand the purchasing of cucumbers. Data from one customer isn't enough, so the proposal is to combine data from all customers into one table and then calculate transition probabilities based on this. Is this a reasonable method? What are some modelling assumptions that have been made? What are the resultant limitations of any analysis that relies on this method? Find transition probabilities for each customer and for an overall set of data from every customer and compare these.

4. There are some products that customers buy almost every week, if they miss a week they always return to purchasing the next week.
   (a) Identify one such product for one customer, making its transition probability matrix to justify your conclusion.
   (b) How would you describe customer behaviour in relation to this product?
   (c) How would you explain this behaviour?
   (d) What action might you propose the supermarket takes to avoid the skipped weeks?
5. There are other products that customers only buy occasionally, then it takes a while for them to return to buy another, and they rarely if ever buy two weeks in a row.
   (a) Identify one such product for one customer, making its transition probability matrix to justify your conclusion.
   (b) How would you describe customer behaviour in relation to this product?
   (c) How would you explain this behaviour?
   (d) What action might you propose the supermarket takes to make them buy it more often?