

1. The data to be analysed relates to purchases made by several customers during the month of December at a small retailer.
  - a. Load the *compras.txt* dataset and perform a preliminary analysis of it:
    - Number of customers
    - Number of different products sold
    - Number of shopping baskets
    - Number of purchases of each product numerically and graphically using a histogram
    - Number of products per shopping basket
    - Count the baskets by number of products sold, that is, the number of baskets with 1 product, with 2 products, ...
  - b. With the data previously loaded create a basket object, with each basket described as a one-hot encoded and visualize it.
  - c. Use the Apriori algorithm to obtain the frequent itemsets with  $\text{Sup}_{\min} = 5\%$ . Visualize them.
  - d. View frequent sets with 2 or more items and  $\text{Sup}_{\min} = 15\%$ .
  - e. Apply the Apriori algorithm to extract Association Rules with  $\text{Conf}_{\min} = 80\%$ . Visualize them.
  - f. For the previously generated ruleset, view:
    - the range of values for the Support, Confidence, Lift, Leverage and Conviction measures
    - the 15 rules with the highest lift
    - rules supported by  $\text{LHS} > 15\%$ , confidence  $> 95\%$  and lift  $> 2.5$
    - rules that include the canned item in the rule consequent
    - rules that include the item fish and fruitvegs in the rule antecedent
  - g. Apply the Apriori algorithm again to extract Association Rules with  $\text{Sup}_{\min} = 1\%$ ,  $\text{Conf}_{\min} = 10\%$  and graphically visualize the Confidence  $\times$  Support (Lift) rules
2. It is intended to create a dataframe with data to develop a model that characterises consumers into three groups according to the products they purchase, for example baskets with:
  - Fruits and vegetables, dairy, fish, and meat => healthy;
  - frozen\_meals, canned, charcuterie => lazy;
  - remaining => others.

For that:

- With basket dataframe define the customer type accordingly with the products purchased.
- With the cliente.txt file, build a dataframe that allows you to develop a model to characterise consumers.