# Tutorial 11

1. A local retailer has a database that stores 10,000 transactions of last summer. After analyzing the data, a data science team has identified the following statistics:

   {battery} appears in 6000 transactions

   {sunscreen} appears in 5000 transactions

   {sandals} appears in 4000 transactions

   {bowls} appears in 2000 transactions

   {battery, sunscreen} appears in 1500 transactions

   {battery, sandals} appears in 1000 transactions

   {battery, bowls} appears in 250 transactions

   {battery, sunscreen, sandals} appears in 600 transactions

   (a) What are the support values of the preceding itemsets?

   (b) Assuming the minimum support is 0.05, which itemsets are considered frequent?

   (c) What are the confidence values of {battery} $\rightarrow$ {sunscreen} and {battery, sunscreen} $\rightarrow$ {sandals}? Which of these two rules is more interesting, i.e. has higher values of confidence?

2. Suppose for three products $A$, $B$ and $C$, support($\{A\}$) = 0.6, support($\{B\}$) = 0.6, confidence($\{B\} \rightarrow \{A\}$) = 0.9 and confidence($\{C\} \rightarrow \{A, B\}$) = 0.5. Compute the following quantities.

   (a) Lift($\{A\} \rightarrow \{B\}$)

   (b) Leverage($\{A\} \rightarrow \{B\}$)

   (c) Confidence($\{A\} \rightarrow \{B\}$)

   (d) Lift($\{A, B\} \rightarrow \{C\}$)

3. Consider data set `Epub` in the package `arules`.

   (a) Write code to derive the number of 1-itemsets that are frequent when the minimum support is 0.005. Report the most 5 frequent 1-itemsets.

   (b) With the same minimum support as above, report the number of 2-itemsets that are frequent.

   (c) Write code to get the association rules using `support = 0.001` and `confidence = 0.3`; and then plot a figure to visualize the top 5 rules with highest lift.