Learning Objectives

Your task for this assignment is to investigate the FP growth algorithm and study the association rules generated from FP growth. This assignment aims to equip you with better understanding of Association Rule Mining that you have learnt in week 9. By accomplishing this assignment, you will be familiar with the following concepts:

- Generating Association Rules from Frequent Itemsets
- Rule Generation and Selection Criteria
- Rule Study

General Instructions

We provide you with a Jupyter Notebook Download Jupyter Notebookthat already contains a dataset together with some code that you will need for your assignment. In the notebook, you will find headers for the separate questions and space for your answers. Please write all answers in the indicated spots. At the end, **export your notebook as an HTML file containing your report and all outputs necessary** to complete this assignment. Submit both, the completed ipynb and the html file. In case you decide to solve the extra task as well, you can either include the photo in your html export (check that it really gets exported!) or submit separately as png or jpg.

The assignment must be submitted to Canvas. It will be run through Turnitin, so make sure that everything you submit has been done by you and is formulated **in your own words**.

Note that we will deduct marks if the solution is not submitted in the correct format. You can only submit html and ipynb files, and image files in case you attempt the extra task.

Part 1

Consider a real-world dataset (provided in the Jupyter notebook), mining Association Rules with FPgrowth algorithm. Answer the following questions using `the provided code indicated as "Part 1"`. For all the questions put the answers in the dedicated answer cells.

A. What association rules can be found in given dataset with the minimum support 60% and the minimum confidence 70%? (The output is sufficient)

B. What association rules can be found in given dataset with the minimum support 60% and the minimum lift value1.2? (The output is sufficient)

C. Examine the differences from the two sets of associations rules above. Select one rule that has low confidence and high lift value. Why is this rule interesting?

D. Describing results: A summary of results (number of rules, general description), and a selection of rules that you deem is interesting based on “interestingness" measure(s) of your choice. Describe your reasoning behind any choices you make.

Part 2

You now have to create your own FP growth algorithm. Modify the given program to enable your FP growth algorithm to obtain the same output as Part 1. Run your program `on the provided dataset (the same as in Part 1)` to answer the following questions. (Hint: keep in mind the support count and consistent sorting principle.) You can do tasks (c)-(f) even if you don't solve the first ones.

A. What association rules can be found with `in this part` in given dataset with the minimum support 60% and the minimum confidence 70%? Comparing to the output from Part 1(a), what do you think the cause is that leads to a different outcome?

B. Fix the problem you identified in the code and re-run your algorithm to obtain the same results as in Part 1(a).

C. Explain in your own words how the anti-monotonicity property of itemsets is maintained to reduce the search space in generating frequent itemsets. (This and the following are general questions, not code-specific)

D. Explain in your own words how the pattern growth is achieved in generating frequent itemsets.

E. Explain in your own words how many data scans are required for FP growth compared to Apriori.

F. Explain in your own words how time complexity and memory consumption are reduced compared to Apriori.

Extra Marks (more interesting, fewer marks)

Please provide a different item ordering (i.e., not descending order of frequency) in constructing FP-Tree and draw the corresponding FP-Tree of the same dataset shown in Parts 1&2. Do you believe that you have the most compressed (or compact) FP-Tree representation in this particular case? You are welcome to draw with hand and upload a photo to your notebook, but please make sure that your tree is understandable and the photo is of good quality.
If you would like to include the FP Tree in your notebook, you can do so via the command

```
from IPython.display import Image
Image(filename="filename.jpg")
```

Otherwise, please upload the file separately and leave a comment pointing to it in your notebook.