CREATIVE ASSIGNMENT 4
CORE METHODS IN EDUCATIONAL DATA MINING
SEQUENTIAL PATTERN MINING
DUE TUESDAY DECEMBER 8

The goal of this assignment is to find sequential patterns of interest in the data, using a sequential pattern mining algorithm of your choice, using the data in CA4-data.csv

This data set, drawn from an intelligent tutoring system for algebra, has eight variables:
• The lesson in the learning software
• The student
• The skill (given arbitrary numbers)
• If the action was correct (1= correct)
• If the action involved a known misconception (bug) (1=bug)
• If the student requested help (1 = help request)
• How long since the previous student action, in seconds (0 means under half a second)

The goal of this assignment is to find and report back on sequential patterns in the data, which are unlikely to simply be due to chance.

You can conduct sequential pattern mining in any software package, or in Excel, using any sequential pattern mining algorithm you like. You may find it useful to do some quick feature engineering or to create a latent knowledge estimation model. But it's not required.

It's also OK to take a subset of the data, and do analyses on it, if you want. But please justify how you selected the subset of the data, and why.

Please turn in:
• All modified versions of the data file that you create
• All analysis outputs, whether from a software package or Excel
• All data mining code you used to generate the outputs (if relevant)
• A report of what patterns you found, what they might mean, and relevant evidence that they are unlikely to be due to chance. This report should also explain how you completed the assignment You will be graded on completeness and comprehensibility of your hand-in, whether you correctly and validly apply the method you choose to this data, and whether the methods you chose fit the requirements of this assignment.

BONUS: The student who succeeds in producing the "most interesting" pattern which is correctly (or reasonably) interpreted from the data will receive the bonus. Please identify which sequential pattern you would like me to consider as most interesting for the bonus, and explain why it is interesting. (Hand-ins which do not do so will not be eligible for the bonus)