Instructions: This assignment covers topics of classification.

Submission Requirements:

In *w5*, this part, you will be asked to walk through methods and calculations **manually** or display understanding of concepts and topics. In *a5-R* or *a5-python*, you will use packages and libraries to implement classification models, data reduction and text mining methods.

You must prepare your solution to this part as a document using LaTeX. I have provided a template for this assignment, where you can create your answers.

For this assignment, you are to work in your **groups**. I highly suggest using Overleaf to work on your submission together in your group.

Follow the submission template where work for each question must **start on a new page**. Do not put work for multiple problems on the same page.

Questions:

1. Association Analysis I
   Given a database of transactions and a min-support = 2,

Trans.	Items
T1	A, B, C, D, E, F, G, H, J, K
T2	A, B, C, D, E, F, G, H
T3	A, B, C, D, E
T4	F, G, H
T5	K, P, Q, R, S

   (a) (3 points) How many frequent patterns exist? *Note, you should not list what they are, just provide the number.*

   (b) (4 points) Find an example of an association rule that matches the following pattern with min-support = 2 and min-conf = 0.70,

$$(A, B, C, D, ItemX \rightarrow ItemY).$$

   That is, what are the items replace $ItemX$ and $ItemY$. *There may be more than one correct answer.*

   (c) (15 points) For each of the following association rules, report the support, confidence, lift, leverage, and conviction.
   $A \rightarrow F$
   $B, C \rightarrow K$
   $F, K \rightarrow A, B, C$

TID	Items
T1	B, D, F, G, I, J
T2	C, B, D, G, I, J
T3	D, F, G, H
T4	A, D, J, K
T5	A, B, D, E, G
T6	A, B, D, G, I
T7	A, D, G

Table 1: Transaction Data

2. (30 points) Association Analysis II - Apriori
   Consider the data listed in Table 1. Show the operation of the Apriori algorithm. Show the majors steps: present L_i and C_i for each level i identified. Also, report at the end the frequent itemsets identified. Use a min-support threshold of 3.

   Report C_i and L_i in tables with the support for each itemset. Present itemsets in alphabetical order, e.g., A, D, G, AB, AD, BD, DC, etc. If the algorithm halts report a single row of NAs in the remaining tables.

3. Consider the data listed in Table 1. Run the FP-growth algorithm.
   (a) (18 points) Report the FP-tree. This should include the header table, node links and the tree. See slides 51-62 of 07.association-analysis.part1 slide deck. You should create something like slide 62. *Remember to order your header table; if there are any ties sort alphabetically.*
   (b) (8 points (bonus)) Report the conditional pattern base table. There should be a column for items, conditional pattern, conditional FP-tree, and frequent patterns. See slide 64-73 and 74-81 of 07.association-analysis.part1 slide deck. The table you create should look like slide 81. *Remember to order the addition of items to this conditional pattern base table.*