DMAT – Workbook 3 - Association Rules

Name: Guy-Erne Stanislas Assika Student Number: 3016702

|  |  |
| --- | --- |
| **Course** | MSCBD-DMAT |
| **Stage / Year** | 1 |
| **Module** | Data Mining Algorithms & Techniques |
| **Semester** | 2 |
| **Assignment** | Workbook 3 |
| **Date of Title Issue** | 27th March 2020 |
| **Assignment Deadline** | 31st March 2020 at 23:55 |
| **Assignment Submission** | Upload to Moodle |
| **Assignment Weighting** | 3% of module |

* 4 Question of 20 marks each, answer all questions.
* Edit this document and rename the file as Practical1\_FirstName\_LastName\_StudentNumber.docx
* Add your Name and student number to the top of this document.
* Write your answers into this document using another colour than black.
* Include appropriate screen shots and photos of your handwritten work.
* You will be submitting this document on Moodle. Originality of the work will be verified.

Substantial marks will be awarded for clear presentation including but not limited to

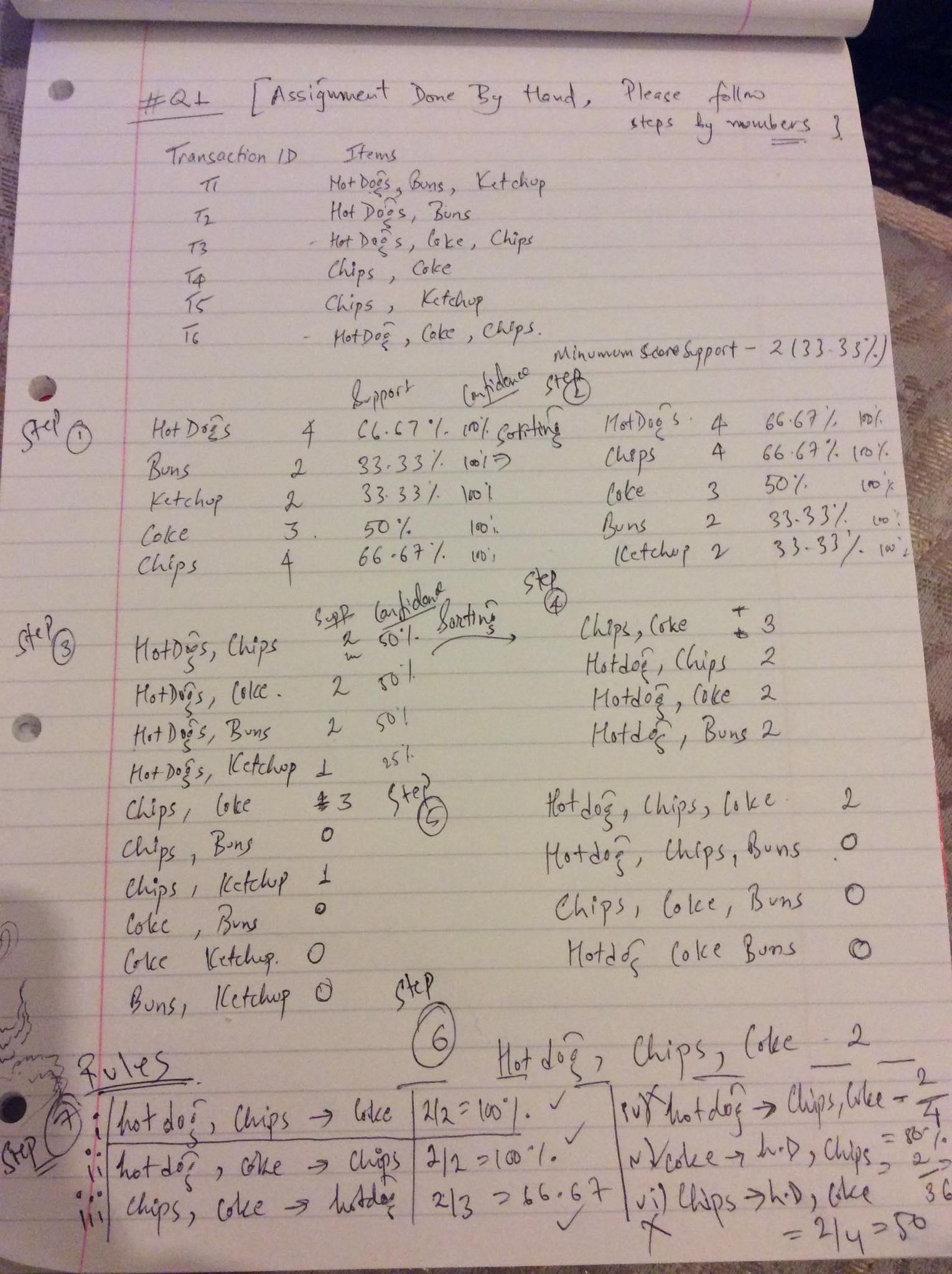
* Correct Question Numbers
* Correct subpart numbering
* Appropriate text density
* Clearly labelled diagrams
* Use appropriately cropped photos for question 1 & 2.

# Q1. By Hand Apriori – 20 marks

1. Illustrate the steps by hand of using the Apriori algorithm on the example in Table A with support threshold *s*=33.33% (2 items) and confidence threshold *c*=60%.
   1. Show the candidate and frequent itemsets for each database scan.
   2. Enumerate all the final frequent itemsets.
   3. Indicate the association rules that are generated
   4. Highlight the strong ones
   5. Sort them by confidence.

|  |  |
| --- | --- |
| Transaction ID | Items |
| T1 | HotDogs, Buns, Ketchup |
| T2 | HotDogs, Buns |
| T3 | HotDogs, Coke, Chips |
| T4 | Chips, Coke |
| T5 | Chips, Ketchup |
| T6 | HotDogs, Coke, Chips |

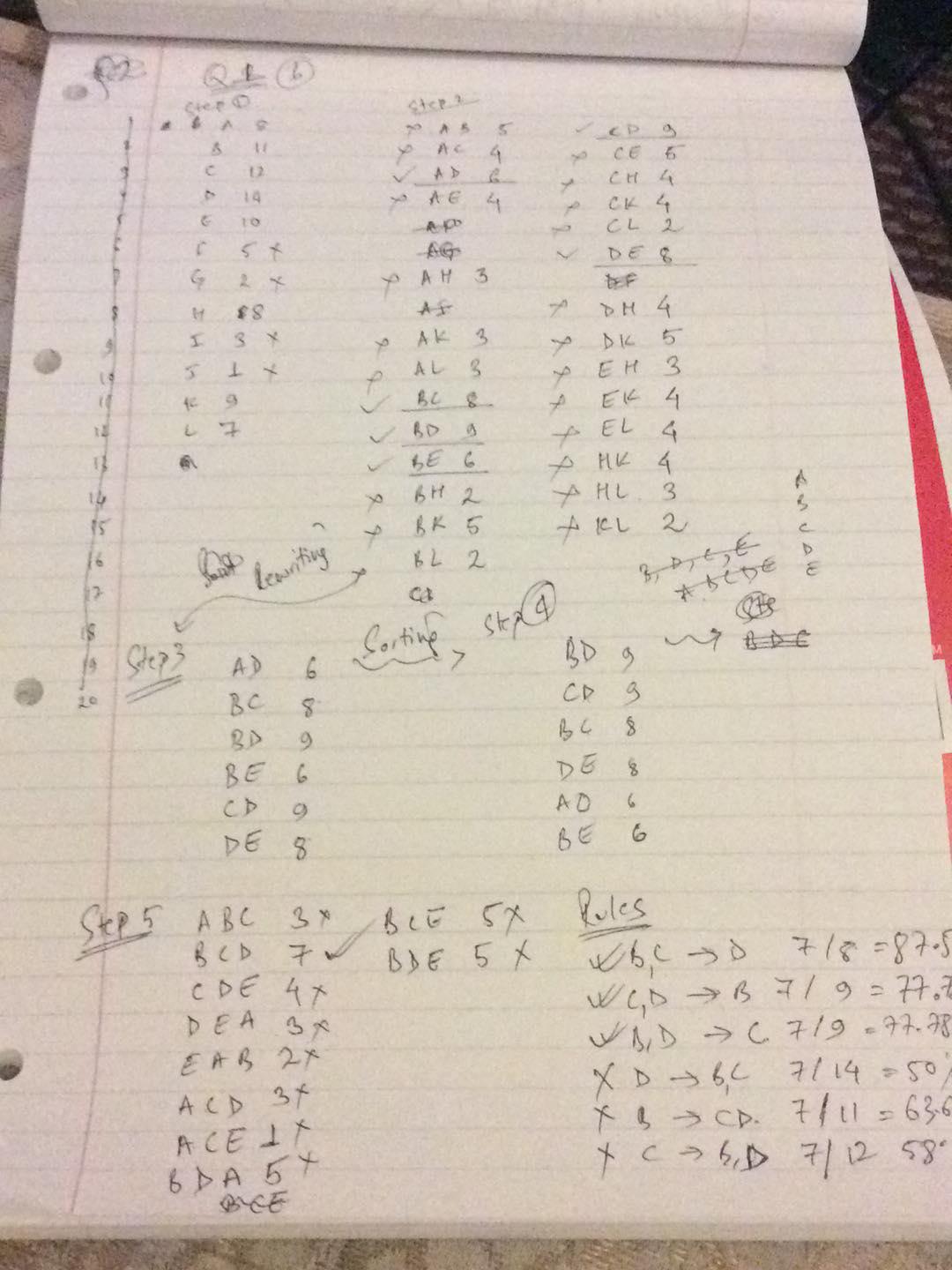
Table A



1. Illustrate the steps by hand of using the Apriori algorithm on the example in Table B with support threshold *s*=30% (6 items) and confidence threshold *c*=75%.
2. Show the candidate and frequent itemsets for each database scan.
3. Enumerate all the final frequent itemsets.
4. Indicate the association rules that are generated
5. Highlight the strong ones
6. Sort them by confidence.

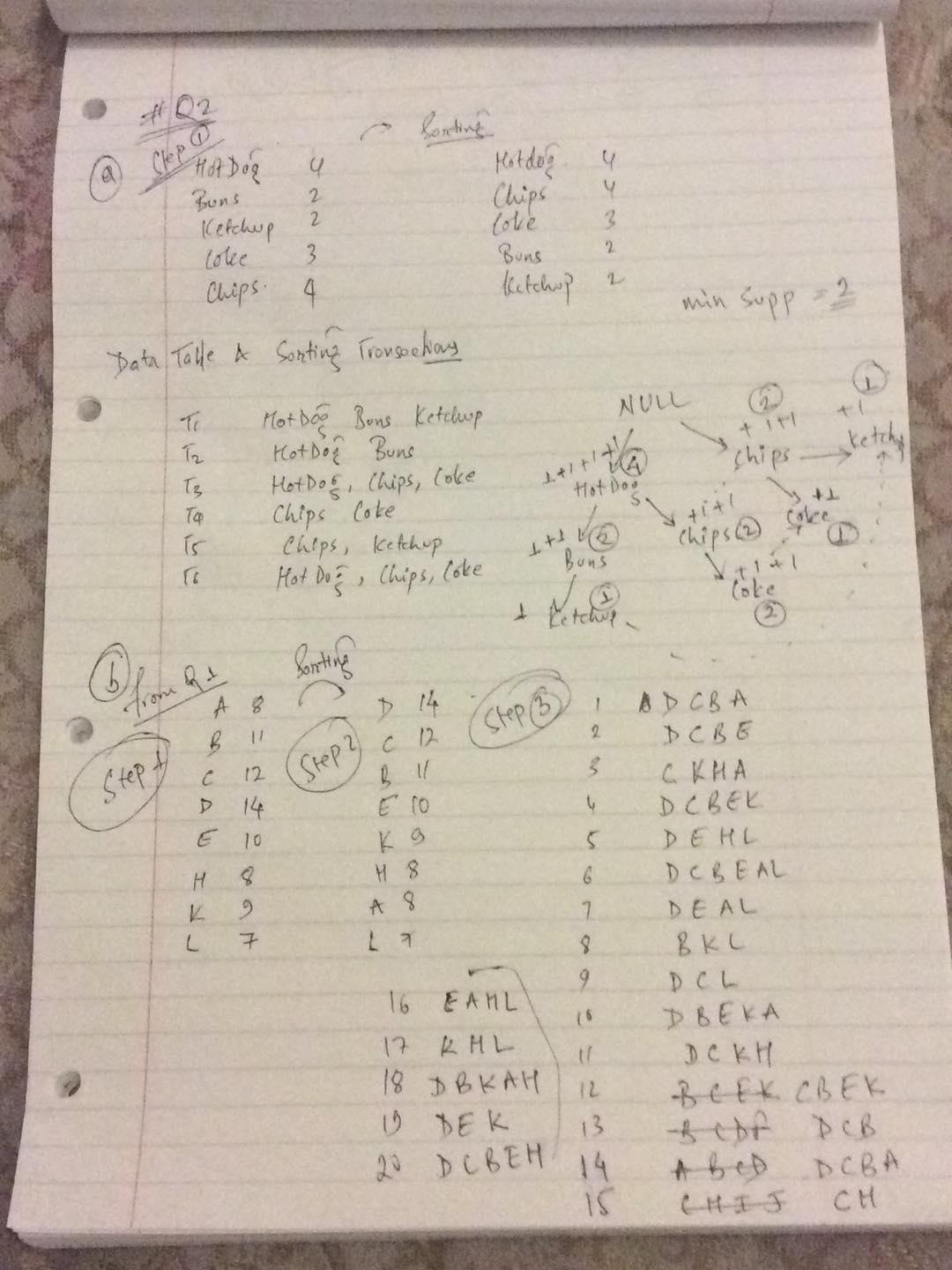
|  |  |
| --- | --- |
| Transaction ID | Items Purchased |
| 1 | A,B,C,D |
| 2 | B,C,D,E,G |
| 3 | A,C,G,H,K |
| 4 | B,C,D,E,K |
| 5 | D,E,F,H,L |
| 6 | A,B,C,D,E,L |
| 7 | A,D,E,F,L |
| 8 | B,I,K,L |
| 9 | C,D,F,L |
| 10 | A,B,D,E,K |
| 11 | C,D,H,I,K |
| 12 | B,C,E,K |
| 13 | B,C,D,F |
| 14 | A,B,C,D |
| 15 | C,H,I,J |
| 16 | A,E,F,H,L |
| 17 | H,K,L |
| 18 | A,B,D,H,K |
| 19 | D,E,K |
| 20 | B,C,D,E,H |

Table B

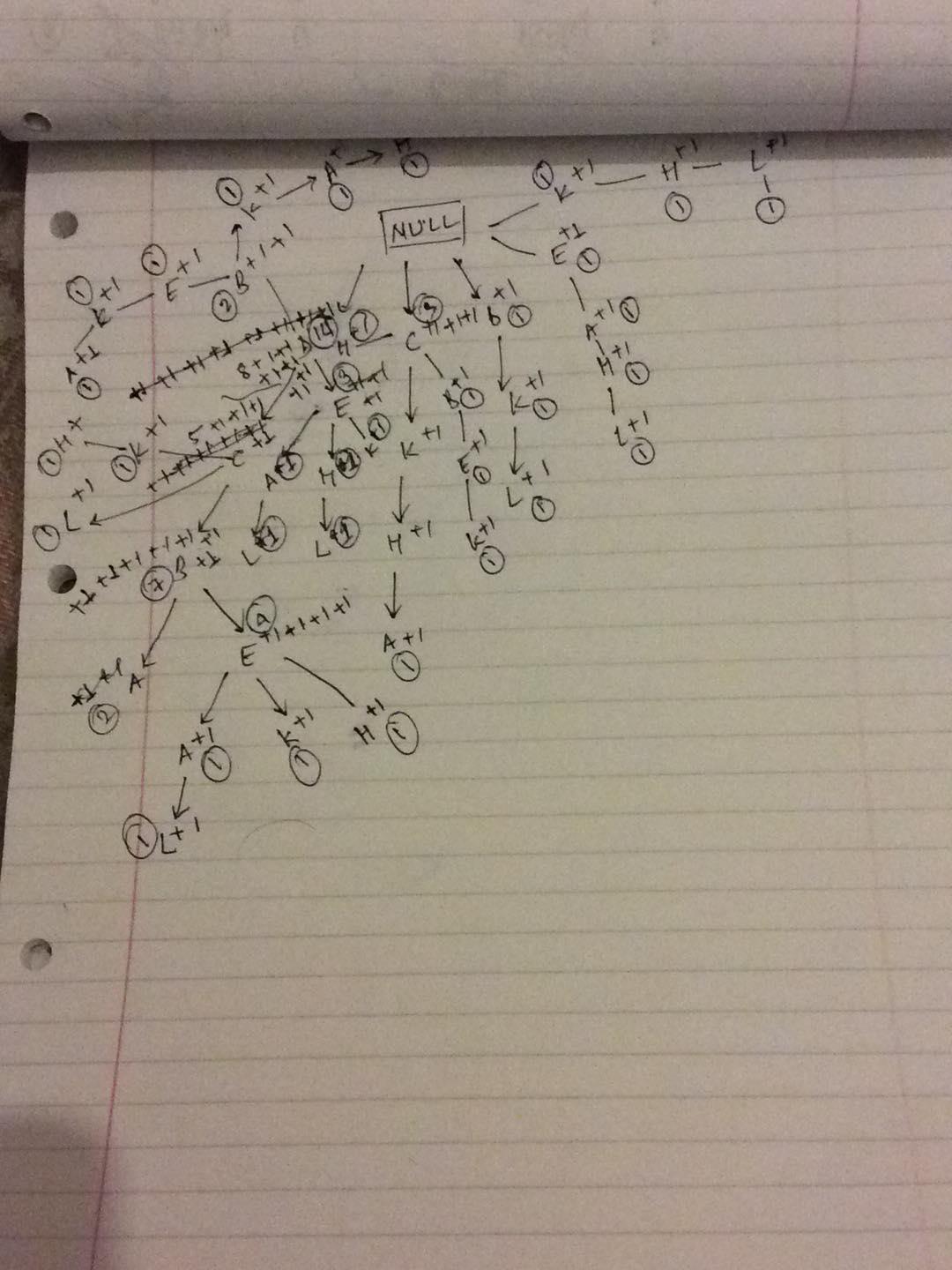


# Q2. By Hand FP-Growth – 20 marks

1. Use the transactional database from Table A in the previous exercise with same support threshold and build a frequent pattern tree (FP-Tree).
2. Show the header table.
3. Show for each transaction how the tree evolves.
4. List the frequent itemsets.



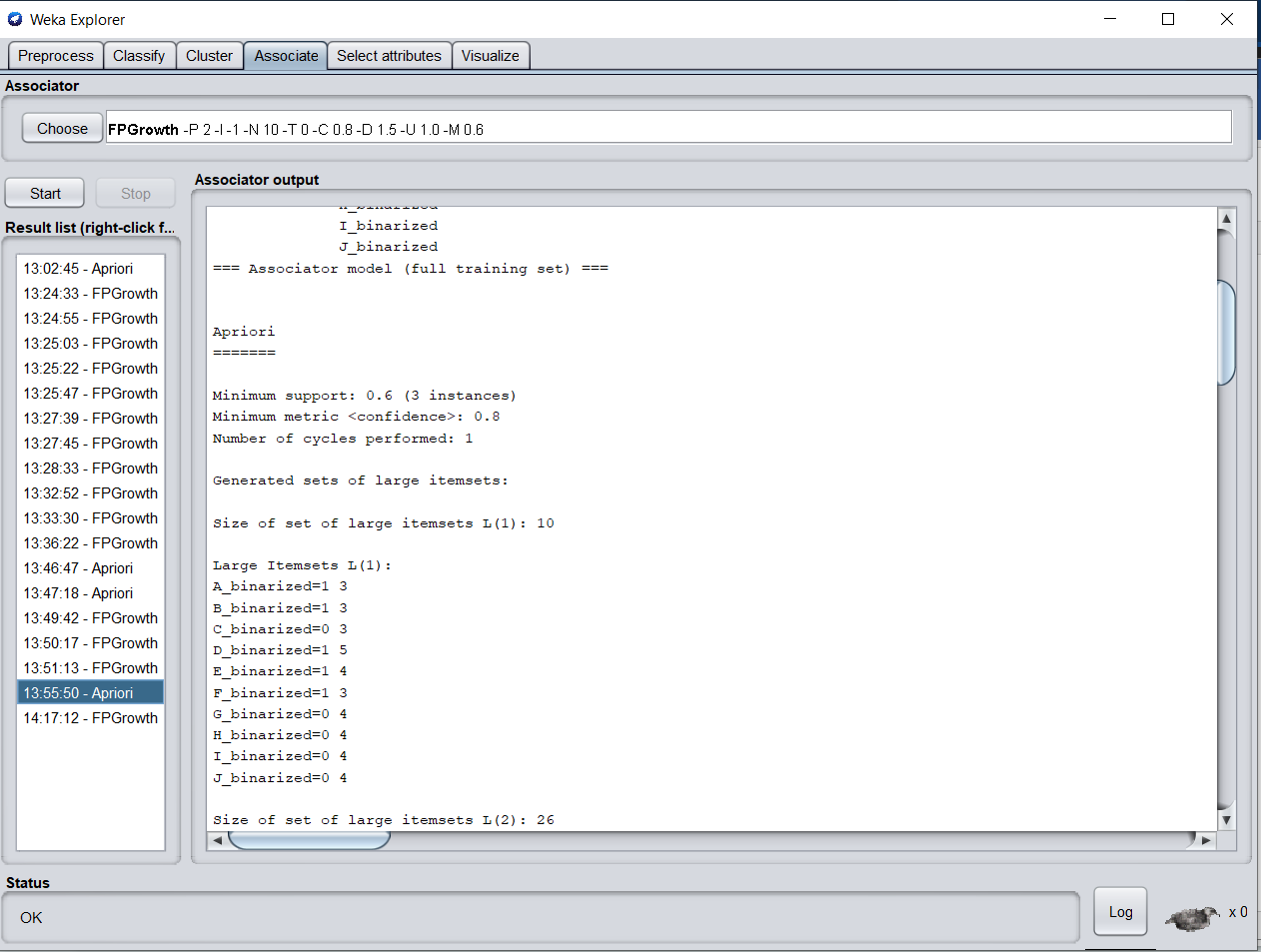
1. Use the transactional database from Table B in the previous exercise with same support threshold and build a frequent pattern tree (FP-Tree).
2. Show the header table.
3. Show how the tree evolves.
4. List the frequent itemsets



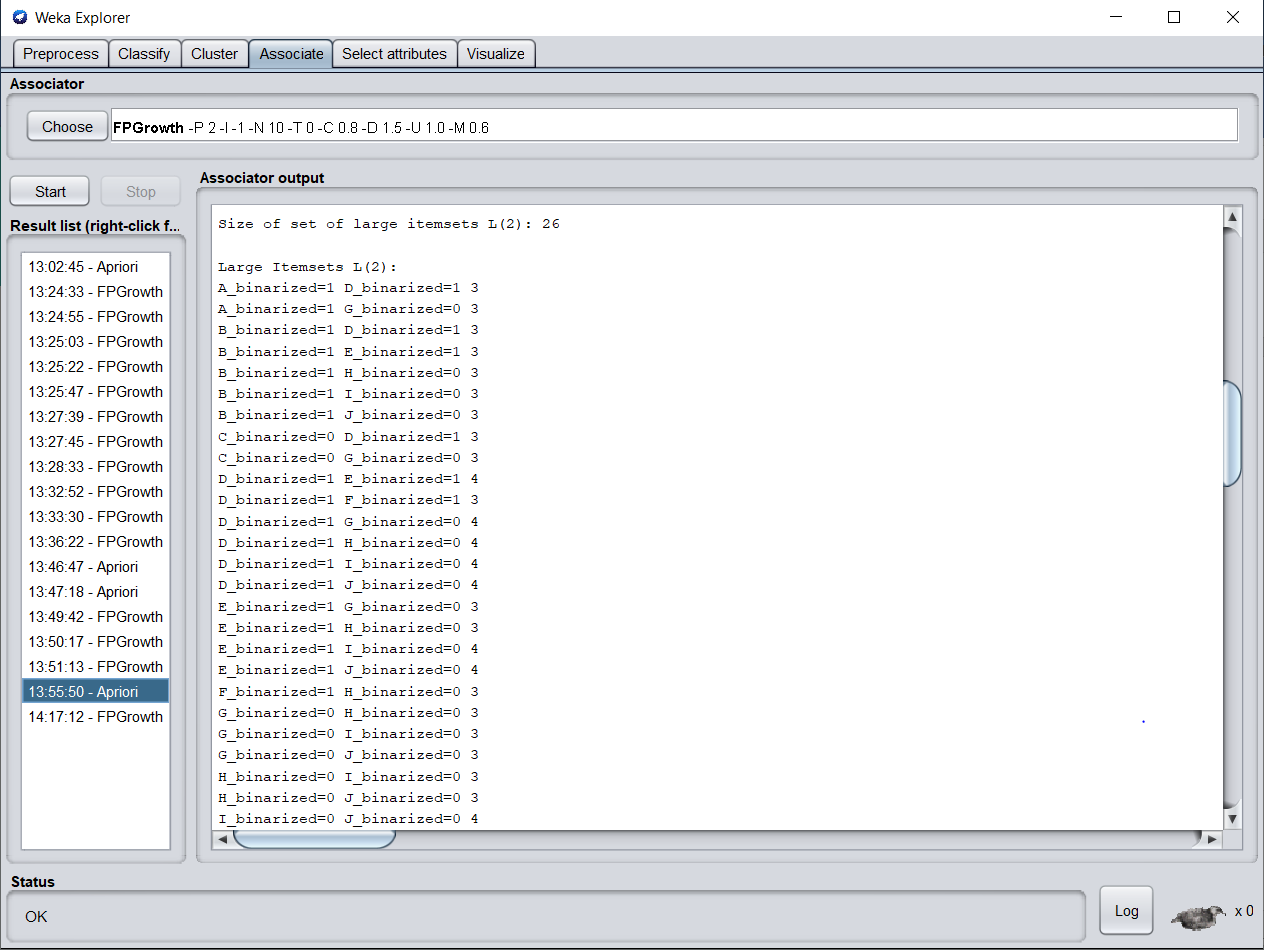
# Q3. Weka – Aprior & FP-Growth – 20 marks

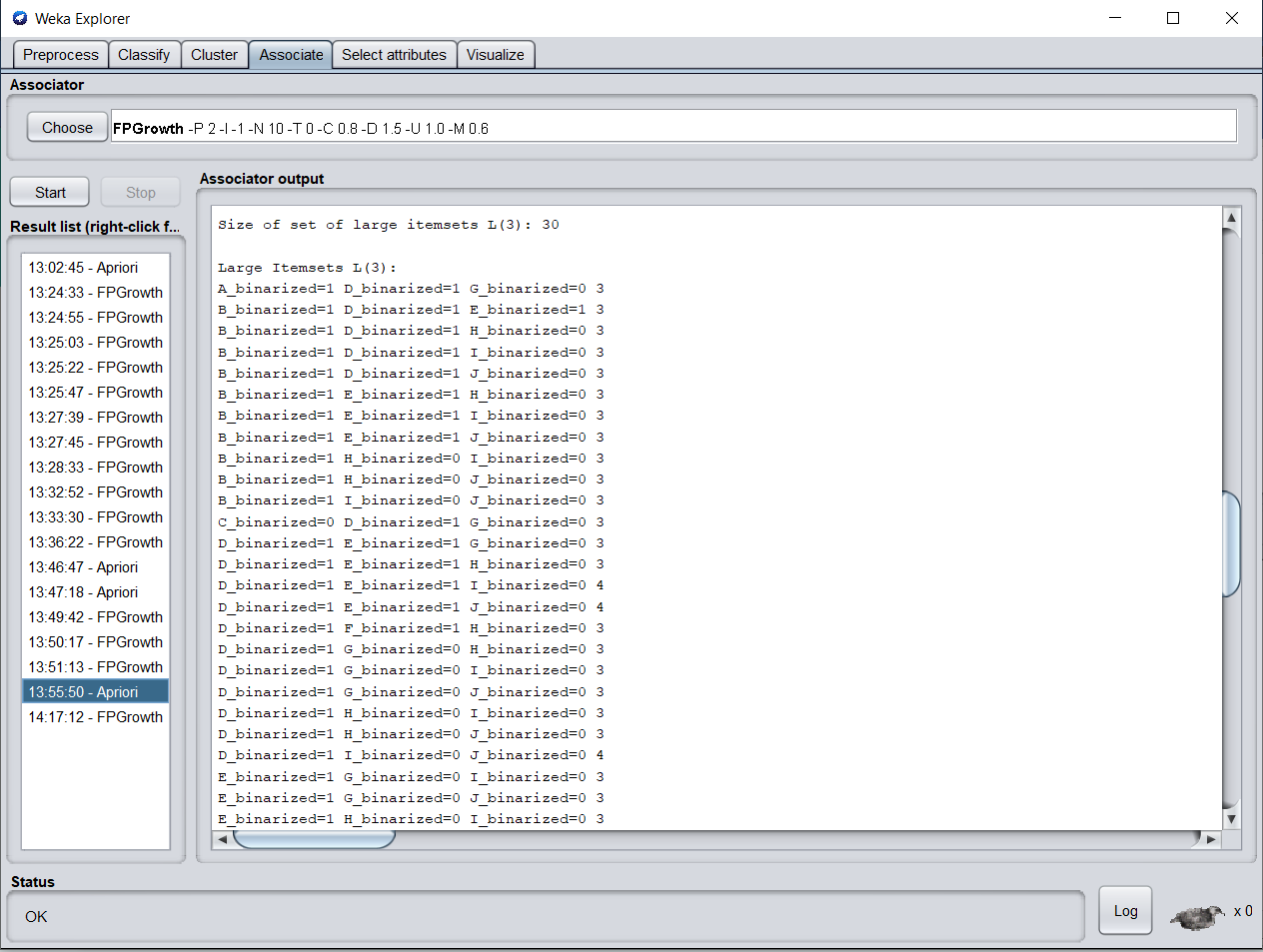
Given the following database in Table C with 5 transactions and a minimum support threshold of 60% and a minimum confidence threshold of 80%. **N.B.** you will have to create the csv file for Table C. Please include this in your submission.

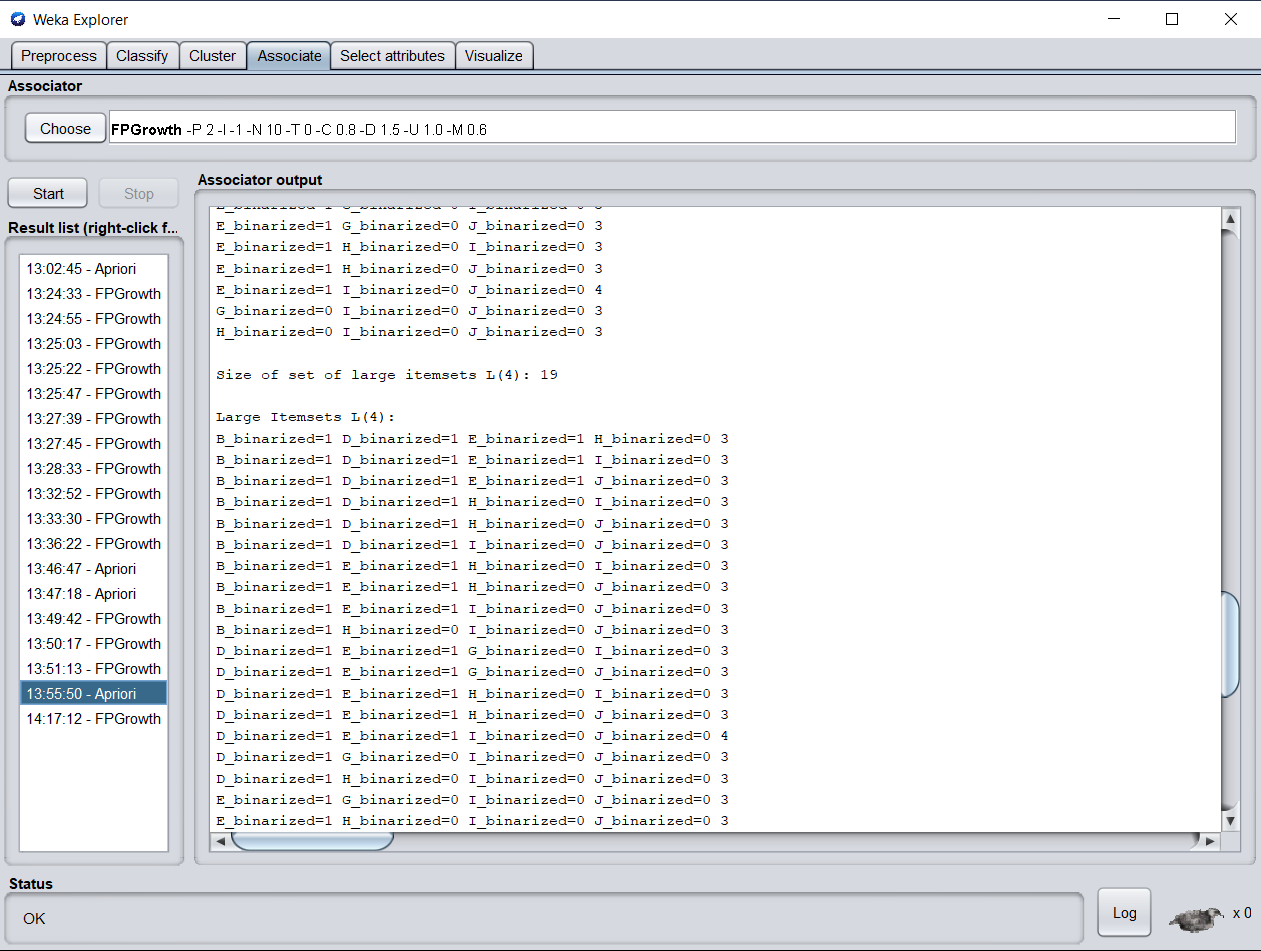
1. Using Apriori with Weka show
2. Itemsets (screen shot)

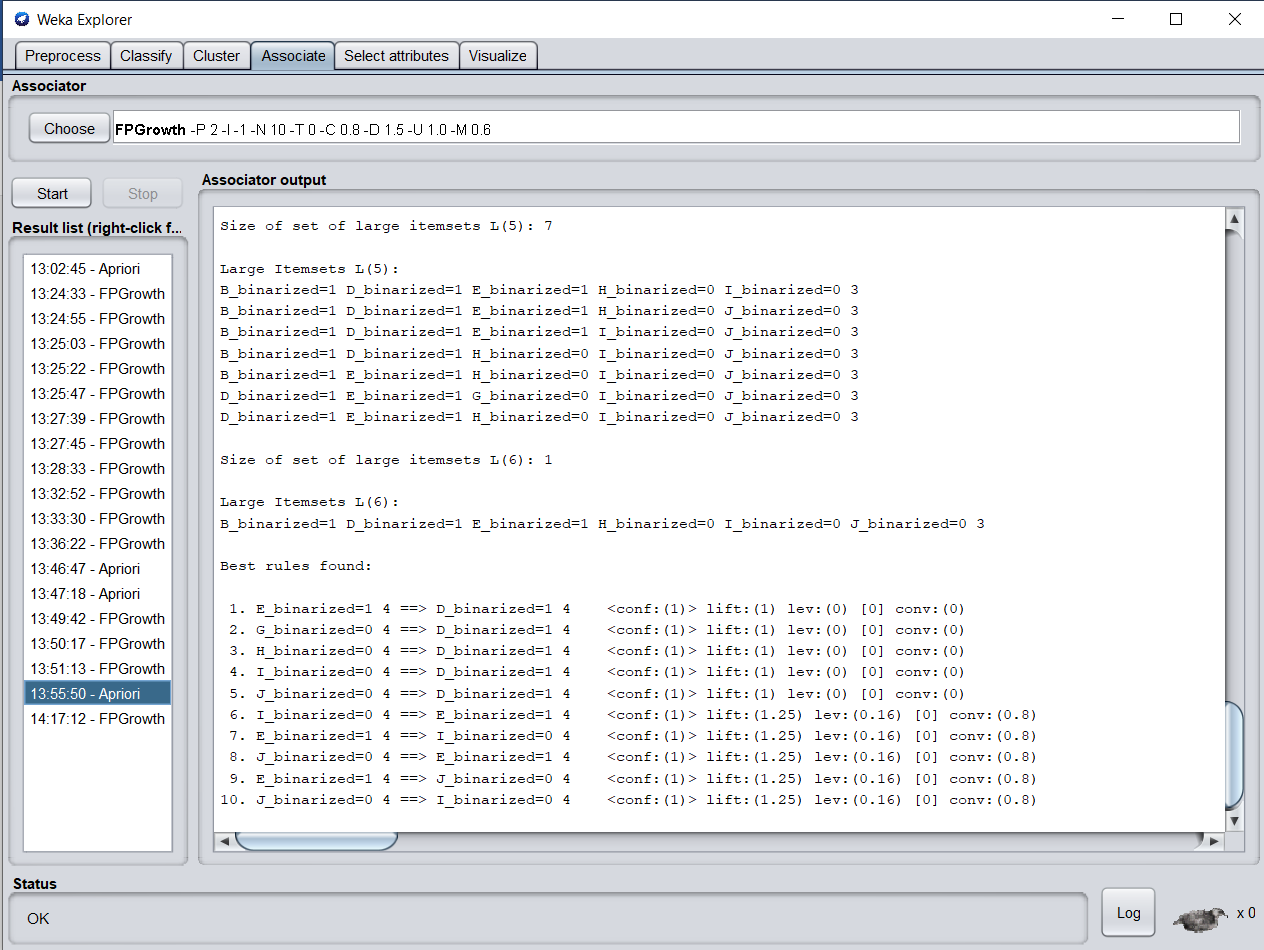


1. Association Rules (screen shot)

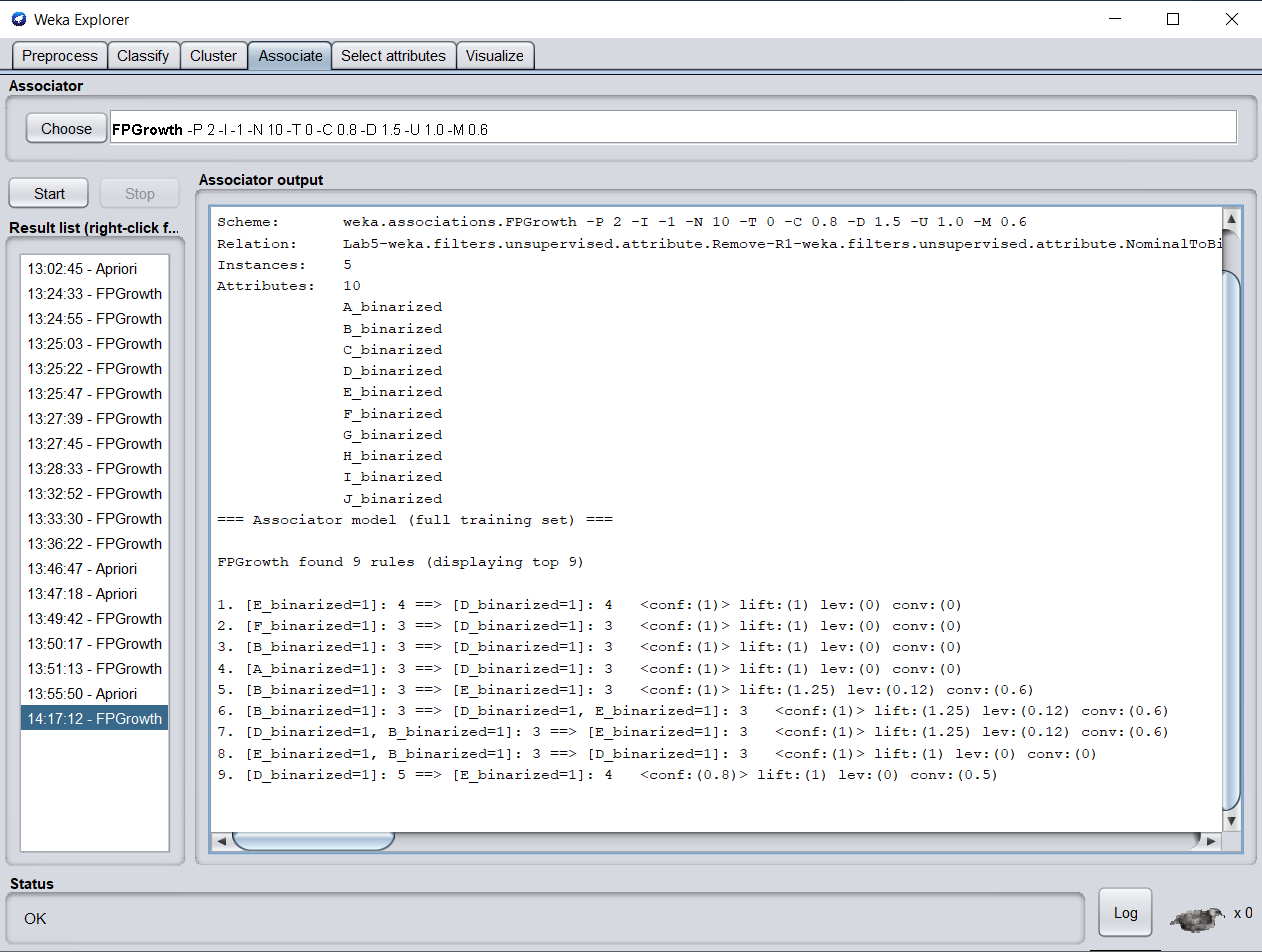








1. Using FP-Growth with Weka show
2. Association Rules (screen shot)



1. A list all strong association rules that contain “A” in the antecedent (Constraint).

Unfortunately, Weka does not provide FP-Tree visualization

|  |  |
| --- | --- |
| **TID** | **Transaction** |
| T1 | A, B, C, D, E, F |
| T2 | B, C, D, E, F, G |
| T3 | A, D, E, H |
| T4 | A, D, F, I, J |
| T5 | B, D, E, K |

Table C

# Q4. Comparison – 20 mark

Using the lecture materials and any other reliable source to support your answer, compare the efficiency of both processes.