**Question 2**

(c): From results in (a), separate out all strong classification rules, i.e., rules that contain the class attribute (survey answer) on the right-hand-side.

**7. distance=2 sex=1 children=1 pets=1 414 ==> answered=1 378** [**conf:(0.91)**](conf:(0.91))

So when customer’s distance is in-middle, he is male and he does have pets – then we can say with 91 percent confidence that customer would answer the questionnaire.

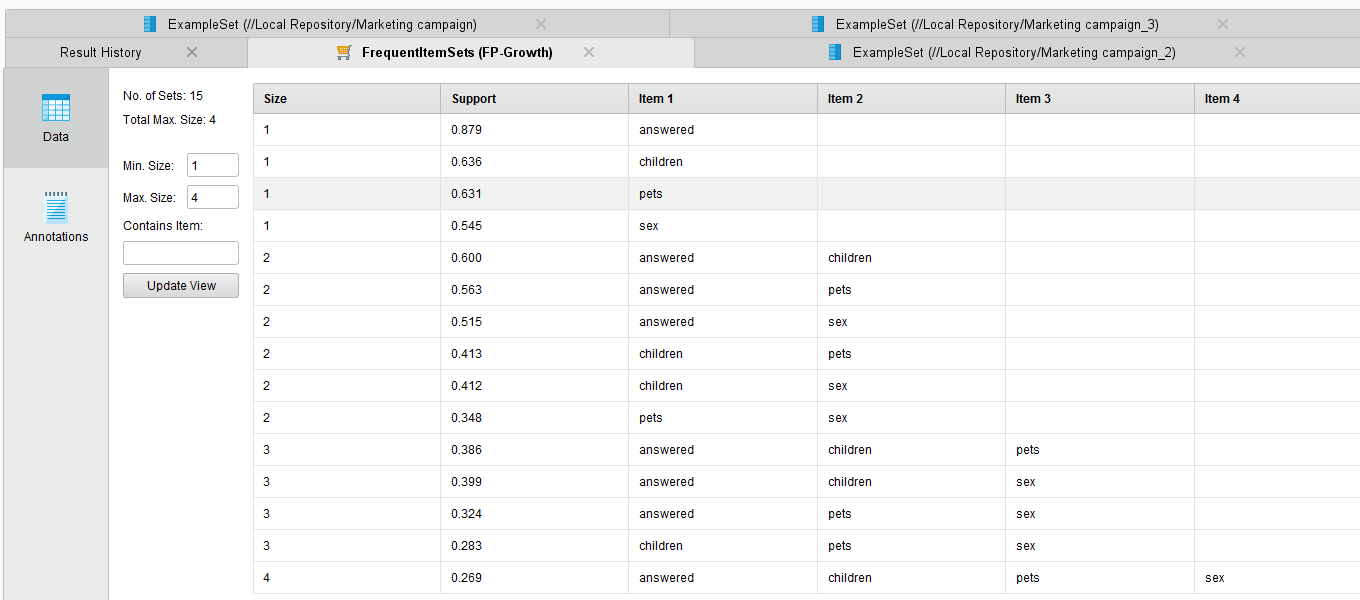
(d): Provide a brief summary and interpretation of results.

**1.** Even though Apriori Algorithm calculates more sets of frequent items (which may include duplicates) but it is beaten by FP-growth Algorithm in that FP Growth has much smaller memory footprint, faster runtime – and is more scalable with huge datasets because of its linear running time.

**2.** For FP-Growth in part (b), a total of 15 tuples were produced that encompassed all possible combinations.

Customers who did not answer the questionnaire had a support of 87.9 percent which was the highest support of all attributes. And as expected, setting min support any higher (e.g. 0.9) produced no results at all.

Also, all four Boolean attributes being false together had a support of 26.9 percent. Checking the “find min number of itemsets ” meant the min support variable was ignored and it decreases support until specified no. of frequent itemsets is found.



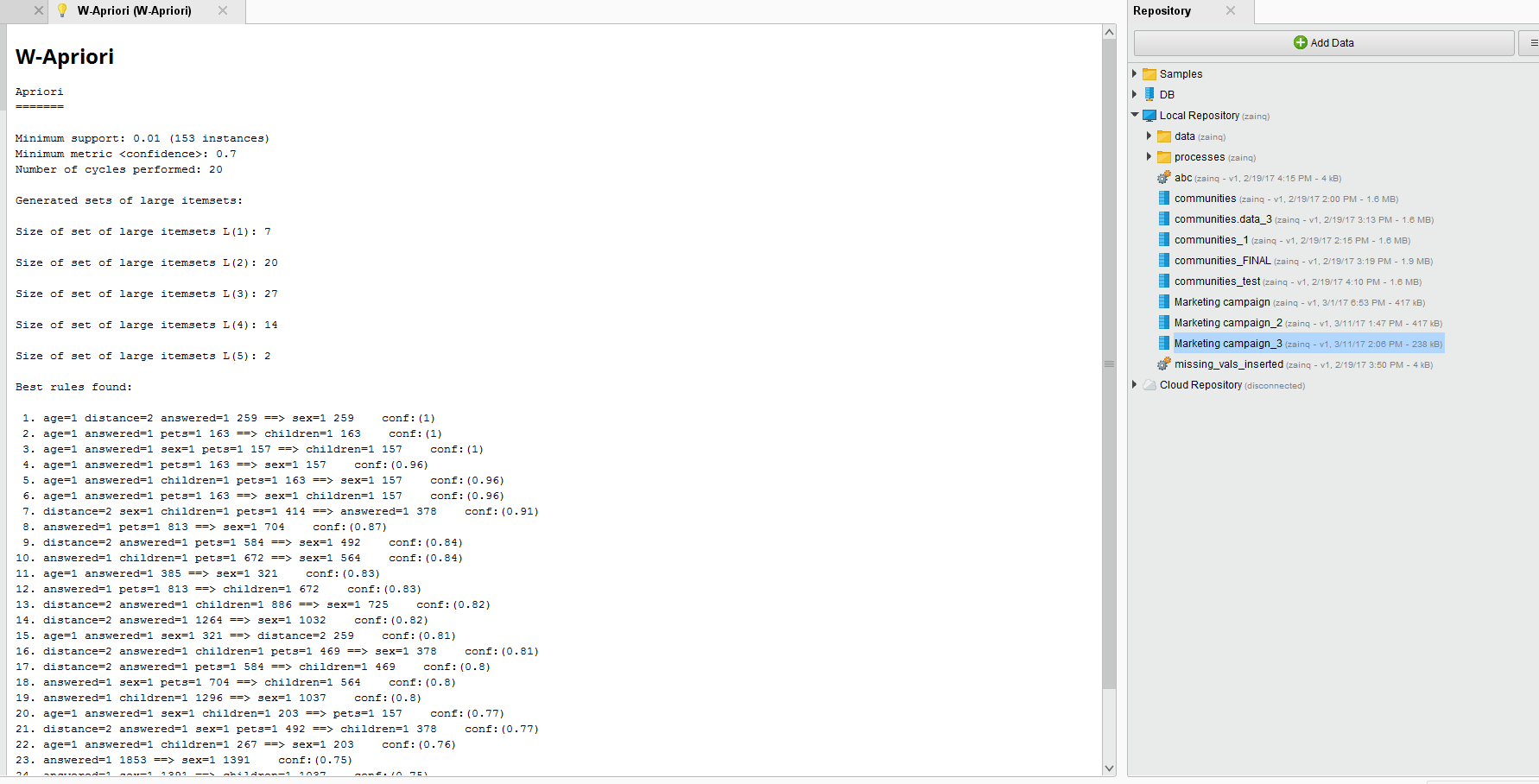
**3.** For Apriori in part (a), one interesting result was when customer’s distance is in-middle, he is male and he does have pets – then we can say with 91 percent confidence that customer would answer the questionnaire.

**distance=2 sex=1 children=1 pets=1 414 ==> answered=1 378 conf:(0.91)**

There were three instances when confidence was 100 percent e.g. when customer is young, his distance is in-middle and he did answer the questionnaire, then we could say FOR CERTAIN – customer was a male.

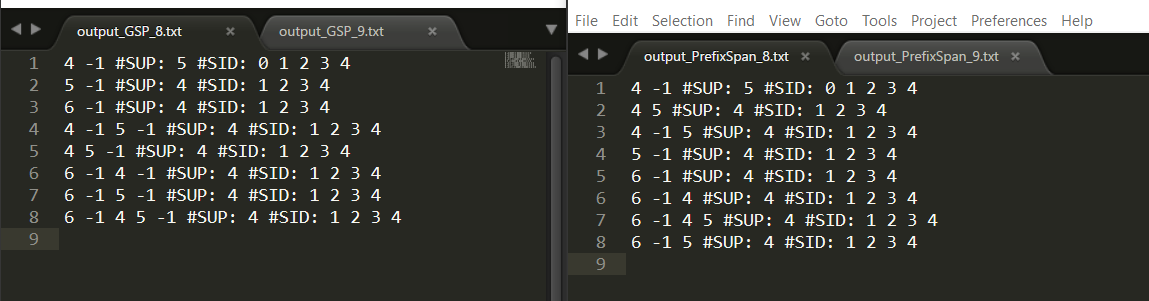
**age=1 distance=2 answered=1 259 ==> sex=1 259** <conf:(1)>

Furthermore, when minimum support is 0.01 then the Apriori algoritm has to perform 20 cycles. This is reduced to 18 cycles when when support is increased to 0.1

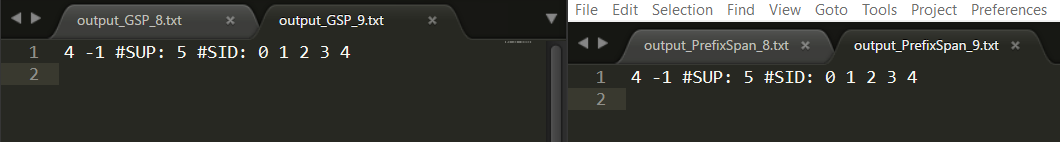


**Question 3**

For at least one minimum support value, verify that the the computed sequences are correct:

**Output of GSP (first) vs PrefixSpan (second) when confidence is 0.8:**

**Output of GSP (first) vs PrefixSpan (second) when confidence is 0.9:**



**Sample screenshot while using PrefixSpan:**

