CS 495/595 –Introduction to Data Mining, HW3

1. (40 pts) Consider the transaction database in the table below:

CustomerID	TransactionID	BasketContent
1	1234	{Aspirin, Panadol}
1	4234	{Aspirin, Sudafed}
2	9373	{Tylenol, Cepacol}
2	9843	{Aspirin, Vitamin C, Sudafed}
3	2941	{Tylenol, Cepacol}
3	2753	{Aspirin, Cepacol}
4	9643	{Aspirin, Vitamin C}
4	9691	{Aspirin, Ibuprofen, Panadol}
5	5313	{Panadol, Vitamin C}
5	1003	{Tylenol, Cepacol, Ibuprofen}
6	5636	{Tylenol, Panadol, Cepacol}
6	3478	{Panadol, Sudafed, Ibuprofen}

   a) For the transactions from the database compute the support and support count for itemsets {Aspirin}, {Tylenol, Cepacol}, {Aspirin, Ibuprofen, Panadol}.
   b) Compute the confidence for the following association rules: {Aspirin, Vitamin C} ⇒ {Sudafed}, {Aspirin} ⇒ {Vitamin C}, {Vitamin C} ⇒ {Aspirin}. Why the results for last two rules are different?

2. Consider the transaction database in the table below:

tid	items
1	a, c, d, e
2	a, d, e, f
3	b, c, d, e, f
4	b, d, e, f
5	b, e, f
6	c, d, e
7	c, e, f
8	d, e, f

   a) Determine the confidence of the rules {a} ⇒ {f}, and {a, e} ⇒ {f} for the transaction database.
   b) Show the candidate itemsets and the frequent itemsets in each level-wise pass of the Apriori algorithm at minimum support of 3.
   c) Construct a FP-Growth Tree for the transaction database at minimum support of 2. Infer the frequent patterns by creating the final result table to generate the conditional pattern base, and conditional-FP-tree.

What to turn in:

Follow the naming convention: **Lastname-hw3.pdf** should contain answers to all above questions. Make sure your name is printed on top of the pdf document.

Submit your pdf file to Blackboard. **Due: April 5 (Sunday) 11.59pm**