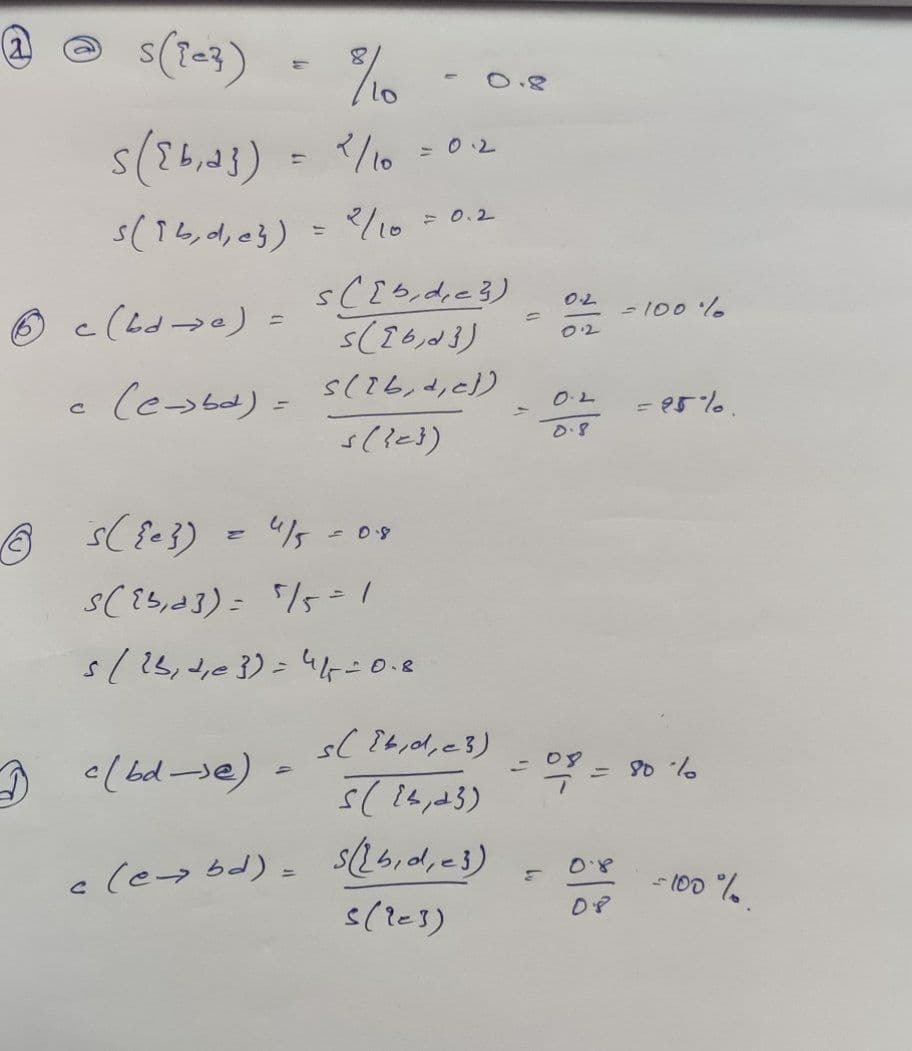
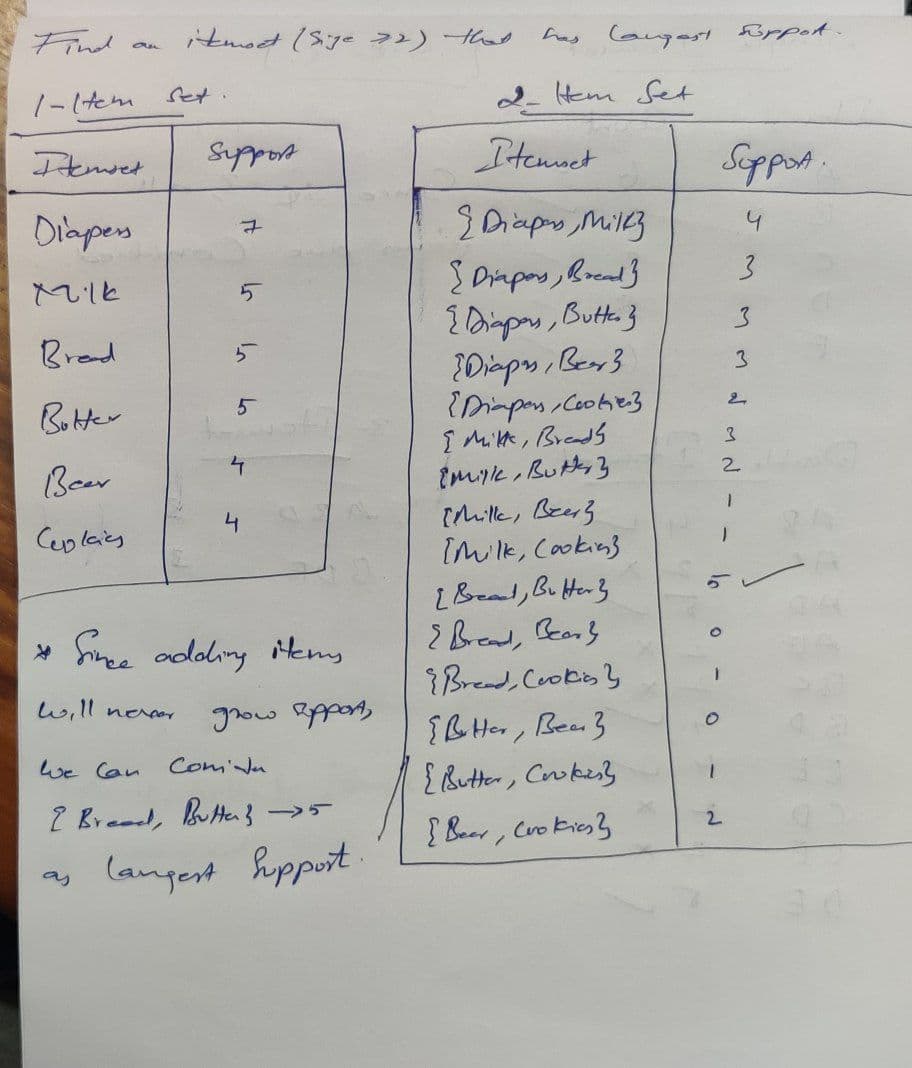
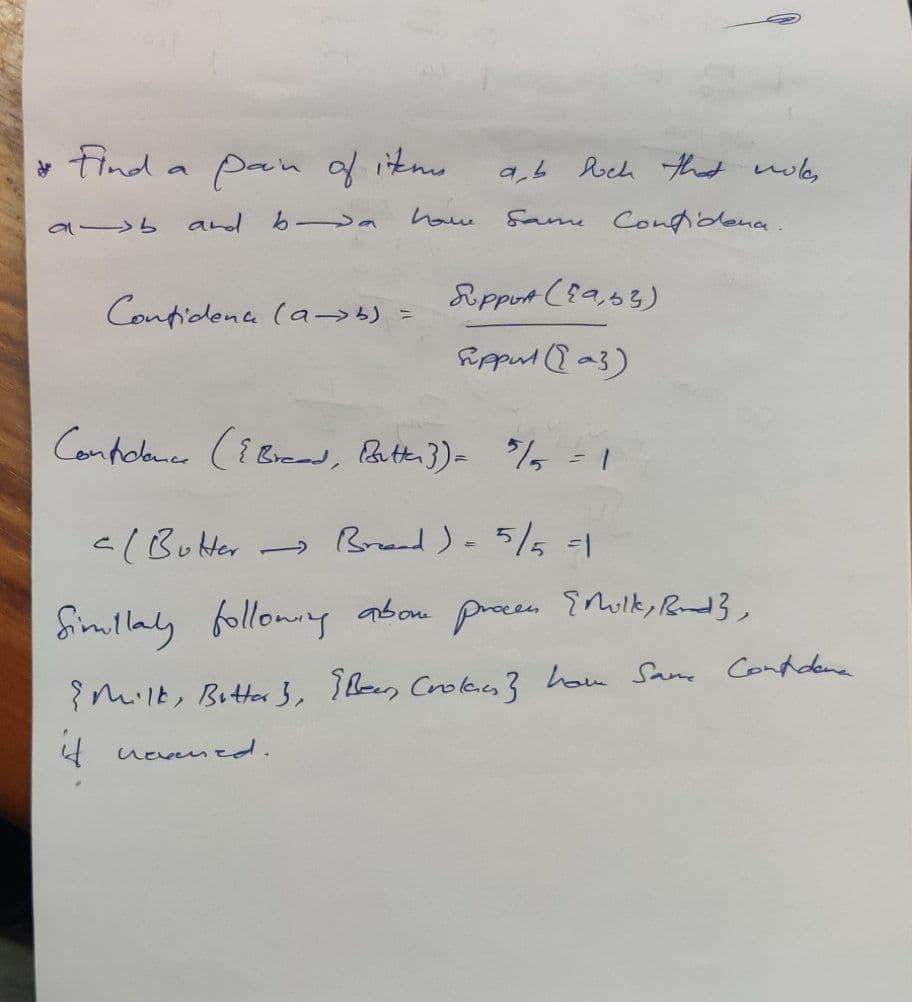
Data Mining Assignment 3

1. Read Chapter 6 (only sections 6.1 and 6.7).  
     
   2) Do Chapter 6 textbook problem #2 (parts a,b,c,d only) on page 404.  
      
   3) Do Chapter 6 textbook problem #6 (parts d,e only) on page 406.  
   



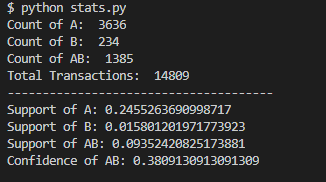
4) Using the data at [www.stats202.com/more\_stats202\_logs.txt](http://www.stats202.com/more_stats202_logs.txt) and treating each row as a "market basket" compute the support and confidence for the rule ip=65.57.245.11 → "Mozilla/5.0 (X11; U; Linux i686 (x86\_64); en-US; rv:1.8.1.3) Gecko/20070309 Firefox/2.0.0.3".

State what the support and confidence values mean in plain English in this context.

Answer:

**Support**: All the transactions that contain all the rules / total number of logs

**Confidence**: Support(all IP rules) / Support(given IP)



**Support**: Total number of transactions that include both items (X) and (Y) calculated as a percentage of total number of transactions. Support gives us a way to calculate how frequently an item of collection of items occur together in terms of percentage of all transactions.

X -> Y

Support = Transactions that contain all items X and Y / Total number of transactions

Here X is a particular given IP address and Y refers to the complete log of information from browser.

**Confidence**: Ratio of number transactions that include all the items from {A} and number of transactions that include all the items from {B} to all the transactions that include items from {A}

X -> Y

Confidence = Support( {X,Y} ) / Support( {X} )

Here X is a particular given IP address and Y refers to the complete log of information from browser