

**part 1**

Support of itemsets

1. 3/10
2. 3/10
3. 2/10
4. 1/10

Rules – support and confidence

5. 3/10    3/7
6. 3/10    3/4

**part 2**

Interesting rules:

1. {administration}  $\Rightarrow$  {bush} The rule administration implies bush has the very high confidence denoting the occurrence of bush is very likely in presence of word administration. The lift is greater than one which shows the high association between the words.
2. {president}  $\Rightarrow$  {bush} This rule has a high confidence of 0.84 and its lift value is 1.36. This implies it's a very good rule, as the likelihood for the occurrence of word bush is much higher in presence of word president.
3. {democratic}  $\Rightarrow$  {primary} The rule democratic implies primary has a very low confidence of 0.60 comparatively and strong lift of 2.1. The low confidence could be due to less occurrence of primary word in general. However, the high lift value shows the word democratic is very likely to appear in presence of democratic.
4. {general, kerry}  $\Rightarrow$  {bush} This rule implies if we have words general and kerry, we are very positive to get the word bush as it has high confidence. The lift and count are also good, but the support is slightly less.
5. {time}  $\Rightarrow$  {bush} This rule has low lift value of 1.07. It shows less association between the words time and bush. Moreover, their confidence is also on the lower side of the average. It is not a good rule relatively.

copy your program code to the bottom

```
cleanDocWords = function(fileName) {  
  # read docword file  
  docWordDf = read.table("docword.kos.txt")  
  colnames(docWordDf) = c("document", "word", "occurrence")  
  
  # read vocab file  
  vocabDf = read.table("vocab.kos.txt")  
  
  # remove file if exists  
  opWordsFile = "words.kos.txt"  
  if(file.exists(opWordsFile)) {  
    file.remove(opWordsFile)  
  }  
  
  for (document in sort(unique(docWordDf$document))) {  
    # get all words in column "word" by document number  
    wordsById = docWordDf[docWordDf$document == document, "word"]  
    words = c()  
  
    for (i in wordsById) {  
      words = c(words, vocabDf[i, ])  
    }  
  
    # write to file  
    cat(words, file=opWordsFile, append=TRUE, sep="\t")  
    cat("\n", file=opWordsFile, append=TRUE)  
  }  
  
  return(opWordsFile)  
}  
  
wordsFileName = cleanDocWords(fileName = "docword.kos.txt")  
t = read.transactions(wordsFileName)  
ruls = apriori(t, parameter = list(support=.2, confidence=.6, minlen=2))  
inspect(ruls)
```