Methodology Summary

1. Filter the data so only the interested data remains
   1. Clean the data (get rid of Date, Year, Month)
   2. Copy all data containing BankA, BankB, BankC, BankD into a new file
      1. One problem with analyzing the data is the runtime, so this will speed up the analyzation
      2. Question #2 of the challenge focuses on if the topic and substance are consistent, so this will help
      3. There are some meaningless content not related to banks
2. Filter the data for meaningless content
   1. Will continue to refine the data we have during each subsequent step
3. Split the data into positive, negative, neutral groups
   1. Use a word list with weight for each word
   2. Will examine data in the neutral group more carefully to check for meaning
4. Sort by topics
   1. Use a word list
   2. Keep word count so frequently used words not in the list may be added to the list
   3. Will consult a professor from Stern
5. Figure out impact
   1. Use a word list
   2. Keep word count so frequently used words not in the list may be added to the list
   3. Will consult a professor from Stern