Project stage 1 report and presentation requirements:

1. Report should be formally organized, formatted, and written without spelling errors.
2. Report and presentation should be printed out and submitted to me before the due day class.
3. Report and presentation should address the following:
    1) What libraries and APIs are used, their websites, and an introduction to them.
        o Provide references at the end of report
    2) Review the key parts of your code to show how your code work on removing html tags, words on the stop list, stemming, and any special handlings for better results.
    3) Any problems you have encountered and solved, and lessons you have learned.
    4) Results:
        o How many words are identified from the training set
        o Present and discuss top 200 most frequent words identified
            ▪ Example words that are not properly cleaned, e.g. sle"), il'-10, etc.
            ▪ Why were they difficult to get properly preprocessed?
        o How many words are identified from the test set
            ▪ How many words appear in both the training set and the test set
            ▪ How many do not?
    5) Note: discussion on lessons and problems won't impact your marks negatively but positively
    6) The presentation slides should be readable to audiences with proper fonts