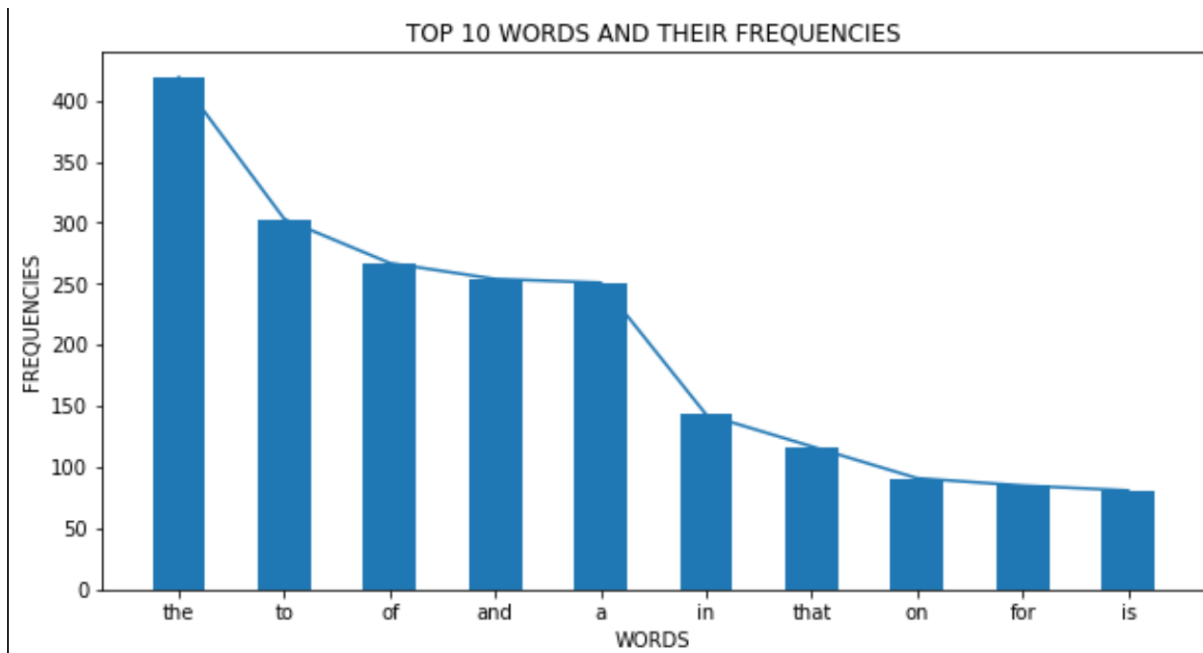
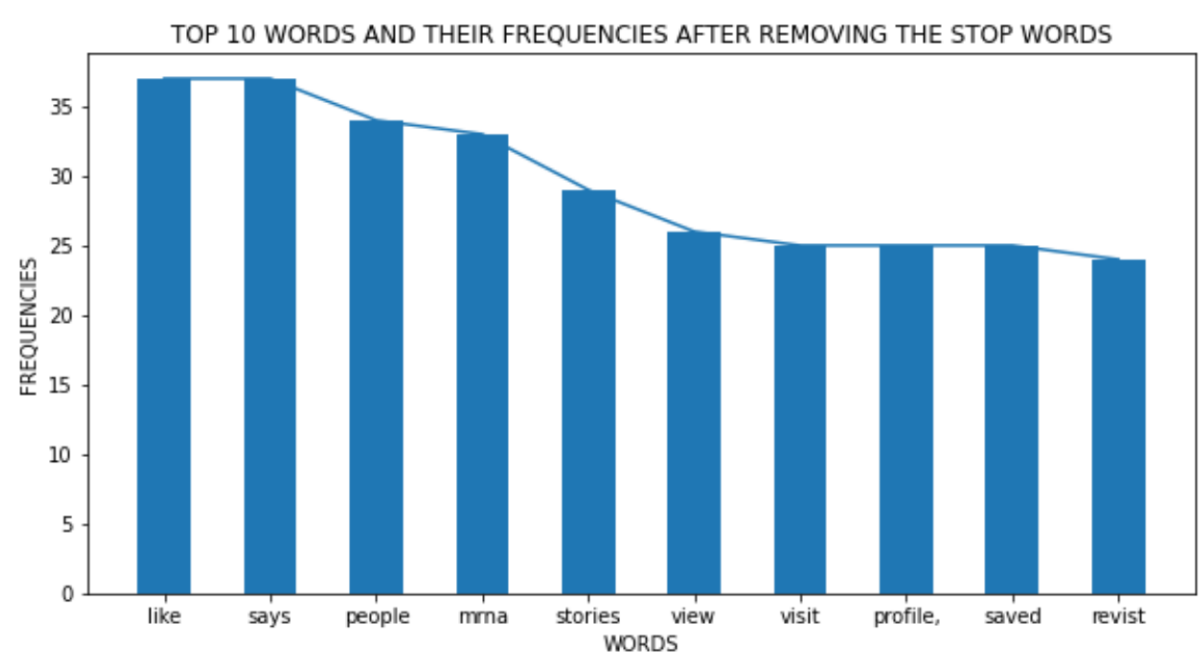


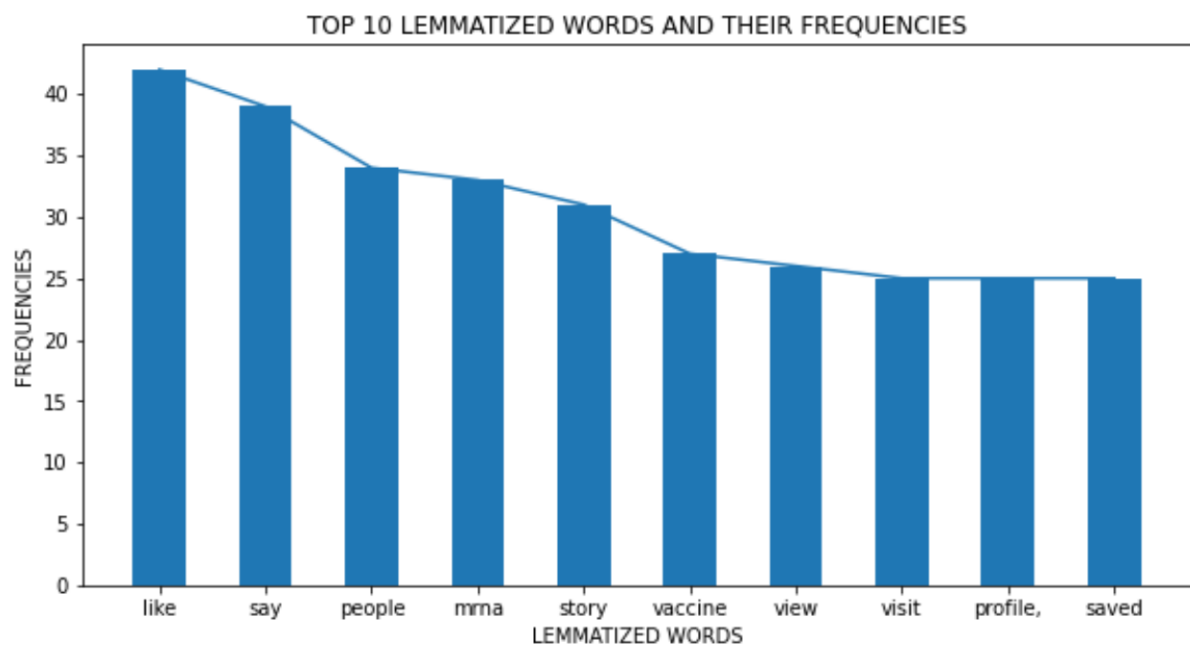
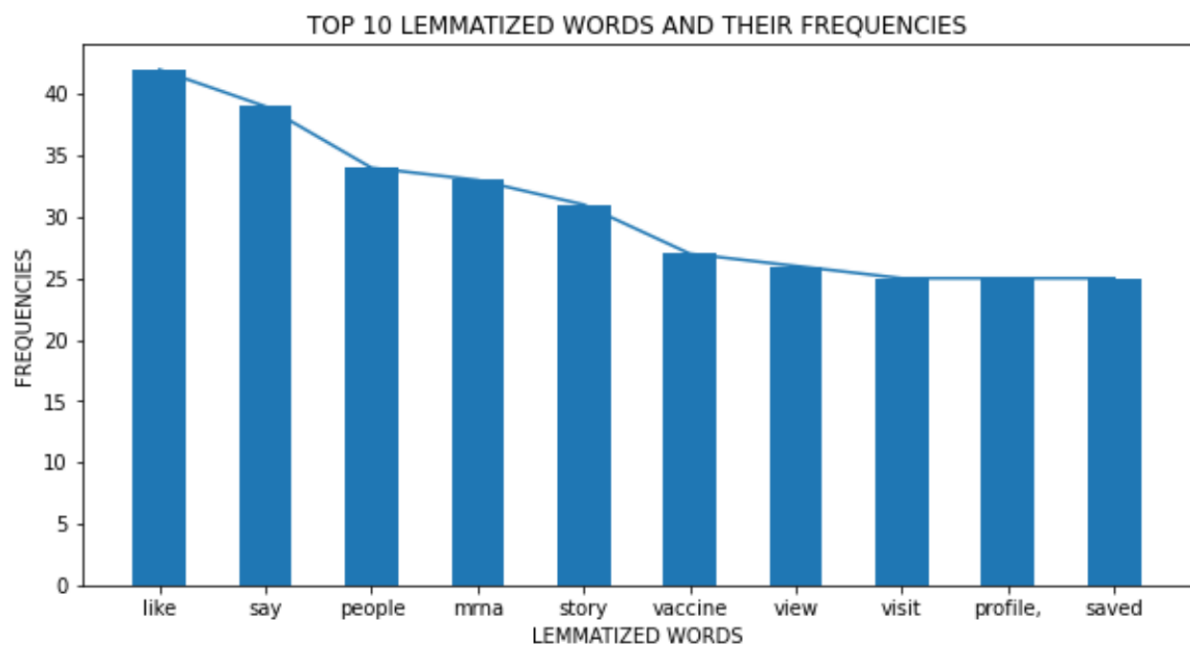
ENGLISH



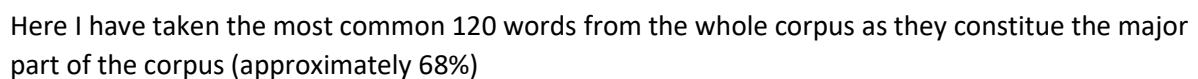
As we can see the majority of the maximum frequency words are stop words in the corpus



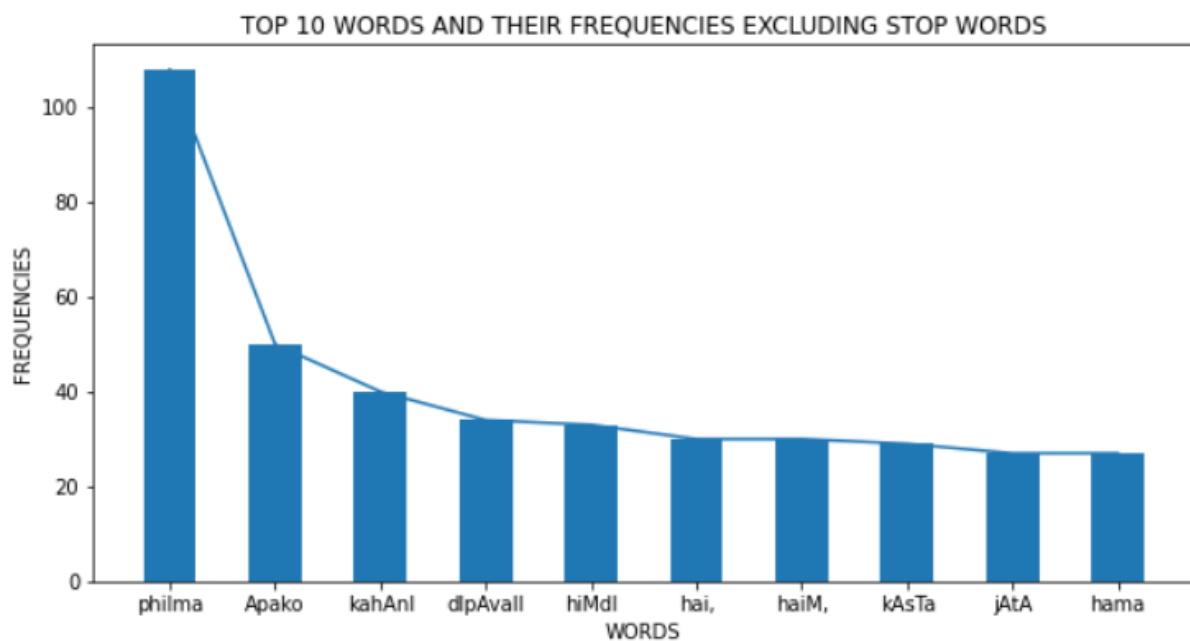
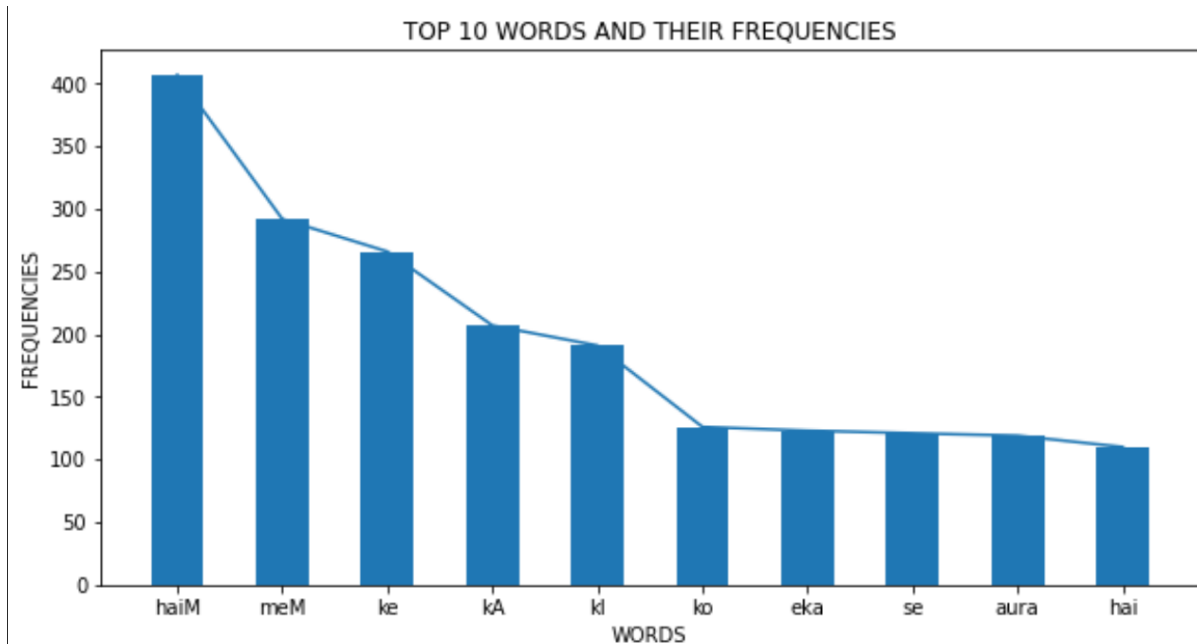
After removing the stop words, we can see the actual real words which have a good frequency in the corp. These words can further help us in understanding what the corpus is about. Therefore, removing stop words is an important part while understanding what a text is about as they just act like noise in the text while doing this type of analysis us



We can see that after lemmatizing and stemming the words, most common occurring words are the same, since lemmatizing and stemming didn't change the numbers by much



HINDI



After removing the stop words, we can see the actual real words which have a good frequency in the corp. These words can further help us in understanding what the corpus is about. Therefore, removing stop words is an important part while understanding what a text is about as they just act like noise in the text while doing this type of analysis us

