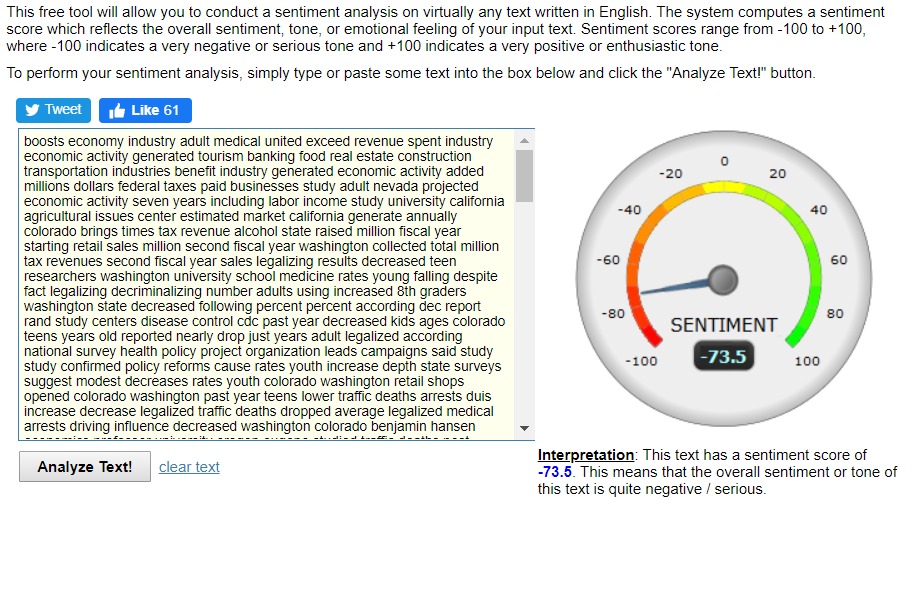
Conclusion:

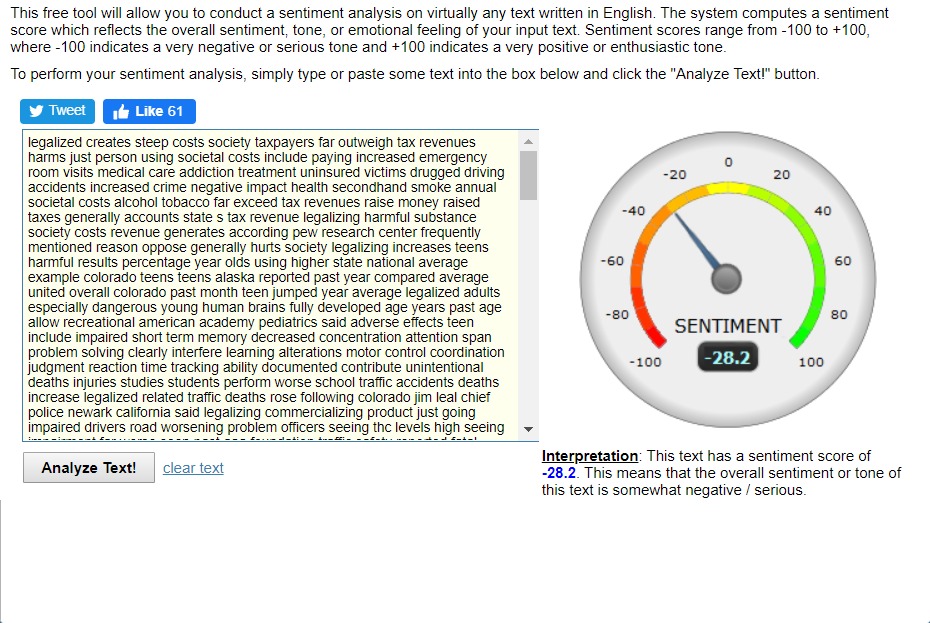
In this exercise I did analysis on two data sets where I removed all stopwords, some non-relevant data like numbers, punctuations, blink lines etc. In addition to that I performed a task to find most frequent words and I appended those words to stopword list and removed those words to improve sentimental analysis. Words like Marijuana, drug, legal, cannabis, legalization, billion, and states etc this were removed to improve the analysis.

When I performed sentimental analyses on website without making any changes in text then it resulted as -51.4 for pro\_ marijuana\_raw.txt file and -22.2 for con\_ marijuana\_raw.txt file.

After performing all the changes i.e. removing all stopwords and punctuations the analysis resulted as -73.5 for pro\_ marijuana\_raw.txt and -28.2 for con\_ marijuana\_raw.txt file.

I am also attaching images of sentimental analysis below for reference.





After all this when I performed word clouds the most frequently used word found where increased, said, users, increasing for pros and Washington, Colorado, alcohol for cons.

By this I can understand that pros have lot more negative words compare to cons. As you can identify most frequently occurring words have larger size in word clouds.

I am also attaching images of word clouds below for reference.

Cons



Pros

