* Sample of text we are processing:

*This movie made it into one of my top 10 most awful movies. Horrible. I don’t care if it makes 1 million, 10 M , or 100. There wasn't a continuous minute where there wasn't a fight with one monster or another. There was no chance for any character development, they were too busy running from one sword fight to another. I had no emotional attachment ( except to the big bad machine ## that wanted to destroy them). If you disagree with me, you can send your thoughts to* [*idonotcare@leavemealone.com*](mailto:idonotcare@leavemealone.com)

Write python code to apply the following on above text and return the final text:

1. Removal of special characters and extra spaces and expanding Contractions
2. Use regular expression to simplify digits and emails
3. Tokenization
4. Removal of Stop words
5. Stemming / Lemmatization

Important notes :

* Build a class that has all the functions to do the preprocessing
* Your class/functions should have docstring and proper type hinting

Templet:

txt\_prep = Text\_preprocessor(string=”here is the string to clen”)

# applying the regular expression

txt\_prep.reguler\_exp(pattern=r”/d+”)

# applying the tokenization

txt\_prep.tokenize(by=’word’)

# removing stop words

Txt\_prep.remove\_stopwords(language=’english’)

# apply stem or lemm

Txt.prep.stem(method=”stem” ) # or method =”lemm”

# get the preproceed text as list

Txt.prep.get\_preprocessed\_text()

>> [here,string,clean]