One of the first steps in working with text data is to pre-process it. It is an essential step before the data is ready for analysis. Majority of available text data is highly unstructured and noisy in nature – to achieve better insights or to build better algorithms, it is necessary to play with clean data. This python code is written to clean the tweets that have been extracted using twitter api. Cleaning includes removing the unnecessary hyperlinks, unwanted symbols and emoticons, e**scaping HTML characters, removal of punctuations marks like question marks and extra dots. The input to this code is the CSV file of annotated tweets. Output will be two text files: one to store the annotations or class to which tweet belong (trending or non-trending**) and the other text file contain the cleaned tweets in lower case.