-: **SENTIMENT ANALYSIS:-**

First, I connected python code to MySQL to import the dataset and insert the outputs using database table.

In data crawling, I performed these tasks:-

* First got the data from input.xlsx using pandas data frame and then inserted it into MySQL table.
* Extracted the URLs from data base and using those URLs, urllib and BeautifulSoup library, I extracted the data from the website by searching for first title and then headings and paragraphs(if exists).
* Opened new file with name Url\_id.txt and wrote html data in the file and closed the file.

#Note:- Actually, while opening these URLs the python ide was showing 406 unacceptable error that’s why, I wrote the code for all but created 6-7 txt files manually to further analyze the text. It was working for all other sites but for these sites, it was throwing error because of Mod\_security.

While sentimental analysis:-

* I opened txt files one by one. Extracted text from those files.
* Tokenize the text to words and sentences.
* Remove punctuation marks from tokenized words
* Remove stop words from the filtered words.
* Used pretrained SentimentIntensityAnalyzer for sentiment analysis. It takes text as input and returns the intensity of various sentiments like positive,negative etc as dictionary.
* Searched for complex words by calculating no of syllables per word while taking special care for es and e as last character(s).
* Calculated FI.
* Calculated no of personal pronouns while taking special care for “Us” and not counting it in no of personal pronouns.
* This way, I calculated all the outputs and then updated them in the database table and then using pandas dataframe, extracted those values to xlsx file.

#Note:- while giving syllable per word as output, I found that it was inconvenient to give syllable per word as output list. So I used average syllable count in the text rather than syllable per word.