Initially, I downloaded the input.csv file and uploaded it on data extraction file. Initially, I tested beautifulsoup to extract article title and article text on a URL link number 2. After testing it out, I saved the article title and article text in a text file and downloaded it. To perform the same on all the URL links, I created a for loop in which I have added the lines to extract the article title and article text using beautifulsoup and I have saved all those extracted texts in a list called ‘content’. This can be seen in data extraction file. Then I downloaded each of the element in the list into separate text files that are numbered according to their URL\_ID. Then I merged all the stopwords into one file and I had to create the positive dictionary and negative dictionary so I matched it with the stopword list and downloaded a revised set of positive dictionary and negative dictionary.

Coming to the data analysis part, first of all I had to execute all the functions computing a variable on a certain text, which is the text of article number 2. I created functions for computing each of the variable that was supposed to be computed. Then in the data analysis combined file I have worked on computing the variables for all the article texts with the help of for loop. First of all I had to upload the input.csv file and then create a column for file paths through which I can access each text file. Then I began writing the function to compute each variable. Before that I had to create a function to tokenize the data and remove stopwords according to the stopword list. I had to load the positive word dictionary and negative word dictionary. To calculate the positive score, negative score, polarity score, and subjectivity score, I had to first tokenize the data and then computed those values in their functions itself. Then for computing the rest of the variables I simply sent the original text data to each function where initially it tokenizes the data and then computes the value. At the end of each function, the value computed for each article is saved in its respective list. Then using for loop I had to iterate through all the text files with the help of file path column and save the computed values in its respective list. Then I added each of the lists containing computed values into separate columns in the csv dataframe as shown in this file. Then I removed the filepath column and downloaded the finaloutput file. For performing the data analysis on all the text files I had to use the allstopwords.txt, pw.txt, and nw.txt files. The pw.txt contains positive word dictionary and nw.txt contains negative word dictionary.