

## INSTRUCTIONS

The code.py contains the python code for extracting the texts from urls and performing the text analysis and Output\_data contains the output in an excel file, in the given output data structure.

BeautifulSoup module was used to extract the data from the urls given.

Only the article heading and the article text was extracted from the url.

Out of 114 urls, 3 urls were invalid so I used 'for' loop for each url and then used 'if' command to only consider the valid urls.

I saved the text, urls and url\_id in a list and converted it to dataframe.

Then I performed text analysis and computed different variables by defining function for each variable.

Then the dataframe obtained is converted to excel file using to\_excel() function in python.