

BLACK COFFER TEXT ANALYSIS

Methodology/Approach:

- Used BeautifulSoup Library to extract articles and titles from the URL's.
- The URL's used two types of tags, so I handled both cases to extract data. Try: Exception block was used to handle bad URL's.
- Extracted stop words from text files and removed all stop words from articles.
- Defined a function for each text analysis metric
- Created a text analysis function which computes all metrics in one function.
- Inspected the results.

Instructions:

- The project was built on google colab platform.
- The Black Coffe folder consists of input.xlsx (input file of URLs), Master Dictionary (positive and negative words), Stop words(all stopwords), URL_id.txt (only extracted articles), output.xlsx(final output file)
- Necessary pip install codes and libraries have already been included in the code. Ctrl+F9 to run all blocks of code.