#Importing Libraries

##For Web Scraping

```
from bs4 import BeautifulSoup
import requests
import csv
import pandas as pd
import numpy as np
```

#Web Scraping

###Scraping The Data And Getting All URLs In A LIst

```
urls = []
with open('url_file.csv') as csv_file:
    csv_reader = csv.DictReader(csv_file)
    for row in csv_reader:
      urls.append(row["URL"])
```

###Getting the Content From All URLs

```
context = []

for i in urls:
  response = requests.get(i)
  soup = BeautifulSoup(response.content,'lxml')
  find_div = soup.find('div',class_= "td-post-content")
  if find_div is not None:
    context.append(find_div.text)
  else:
    context.append("Not Available")
```

#Creating The DataFrame

```
data = zip(urls,context)

df = pd.DataFrame(data,columns=["Websites","Context"])
df
```

```
                                              Websites  \
0    https://insights.blackcoffer.com/ai-in-healthc...
1    https://insights.blackcoffer.com/what-if-the-c...
2    https://insights.blackcoffer.com/what-jobs-wil...
3    https://insights.blackcoffer.com/will-machine-...
4    https://insights.blackcoffer.com/will-ai-repla...
..                                                 ...
109  https://insights.blackcoffer.com/blockchain-fo...
110  https://insights.blackcoffer.com/the-future-of...
111  https://insights.blackcoffer.com/big-data-anal...
112  https://insights.blackcoffer.com/business-anal...
```

```
113  https://insights.blackcoffer.com/challenges-an...

                                                 Context
0    \nIntroduction\n“If anything kills over 10 mil...
1    \nHuman minds, a fascination in itself carryin...
2    \nIntroduction\nAI is rapidly evolving in the ...
3    \n“Anything that could give rise to smarter-th...
4    \n“Machine intelligence is the last invention ...
..                                                 ...
109  \nReconciling with the financial realities of ...
110  \nWhat Is an Investment?\nAn investment is a r...
111  \nQuality and affordable healthcare is a visio...
112  \nAnalytics is a statistical scientific proces...
113  \nBig Data\nTo begin with I shall first like t...

[114 rows x 2 columns]
```

###Replacing the \n from the context

```
df['Context'] = df['Context'].replace(r'\n',' ', regex=True)
```

```
df
```

```
                                              Websites  \
0    https://insights.blackcoffer.com/ai-in-healthc...
1    https://insights.blackcoffer.com/what-if-the-c...
2    https://insights.blackcoffer.com/what-jobs-wil...
3    https://insights.blackcoffer.com/will-machine-...
4    https://insights.blackcoffer.com/will-ai-repla...
..                                                 ...
109  https://insights.blackcoffer.com/blockchain-fo...
110  https://insights.blackcoffer.com/the-future-of...
111  https://insights.blackcoffer.com/big-data-anal...
112  https://insights.blackcoffer.com/business-anal...
113  https://insights.blackcoffer.com/challenges-an...

                                                 Context
0     Introduction “If anything kills over 10 milli...
1     Human minds, a fascination in itself carrying...
2     Introduction AI is rapidly evolving in the em...
3     “Anything that could give rise to smarter-tha...
4     “Machine intelligence is the last invention t...
..                                                 ...
109   Reconciling with the financial realities of a...
110   What Is an Investment? An investment is a res...
111   Quality and affordable healthcare is a vision...
112   Analytics is a statistical scientific process...
113   Big Data To begin with I shall first like to ...

[114 rows x 2 columns]
```

#Importing Libraries

For Sentiment Analysis

```
import nltk
nltk.download('stopwords')
nltk.download('punkt')
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

#warnings
import warnings
warnings.filterwarnings("ignore")
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
```

```
stop_words = set(stopwords.words('english'))
print(stop_words)
```

```
{'your', 'into', 'does', "it's", 'an', "you'll", 'needn', 'during',
'but', 'themselves', 'its', 'then', 'now', 'below', 'up', 'most',
'there', 'theirs', 'at', 'by', 'in', 'aren', 'mightn', 'these', 'all',
'shouldn', 'more', 'shan', 'you', 'as', 'no', 'didn', "couldn't",
'he', 'once', 'hadn', 'itself', 'ain', 'are', 'whom', 'under',
'those', 'above', 'were', 'than', 'some', 'hers', 'if', 'when', 'so',
'we', 'few', "hadn't", 'yourself', 'or', 'out', 'with', 'here',
'ourselves', 'very', 'the', "needn't", 'have', 'yours', 'own',
'while', 'was', 'll', "mightn't", 'what', 'that', 'this', 'a', 'each',
'is', 'yourselves', "wasn't", 'down', 'why', 'until', 'before', 'off',
"shouldn't", 'wasn', 'them', 't', 'against', 'over', 'further',
'myself', 'been', "should've", "haven't", 'herself', 'it', 'my',
'how', 'of', 'wouldn', 'y', 'too', "you're", 'both', 'they', 'won',
'again', 'through', 'just', 'which', 'm', 'such', 'ours', 'because',
'o', 'any', 'do', 'she', 'on', 'isn', "shan't", 'ma', 'don', 'having',
"doesn't", "you'd", "you've", 'to', 's', 'our', 'about', 'other',
"didn't", 'am', "aren't", 'his', 'mustn', 'who', "mustn't", 'has',
'being', 'couldn', 'did', "she's", 'will', 'between', "won't", 'him',
'should', "isn't", 'and', 'for', 'doesn', 'i', 'from', "weren't",
'after', 'had', 'd', 're', "hasn't", "wouldn't", 'himself', 'same',
'only', 'nor', "don't", 'hasn', 'her', 'can', 'weren', 'not', 'their',
'be', "that'll", 'haven', 'me', 'where', 've', 'doing'}
```

```
example = df['Context'][0]
tokens = word_tokenize(example)
tokens
```

```
['Introduction',
 '“',
```

```
'If',
'anything',
'kills',
'over',
'10',
'million',
'people',
'in',
'the',
'next',
'few',
'decades',
',',
'it',
'will',
'be',
'a',
'highly',
'infectious',
'virus',
'rather',
'than',
'a',
'war',
'.',
'Not',
'missiles',
'but',
'microbes.',
'”',
'Bill',
'Gates',
'’',
's',
'remarks',
'at',
'a',
'TED',
'conference',
'in',
'2014',
',',
'right',
'after',
'the',
'world',
'had',
'avoided',
'the',
'Ebola',
```

```
'outbreak',
'.',
'When',
'the',
'new',
',',
'unprecedented',
',',
'invisible',
'virus',
'hit',
'us',
',',
'it',
'met',
'an',
'overwhelmed',
'and',
'unprepared',
'healthcare',
'system',
'and',
'oblivious',
'population',
'.',
'This',
'public',
'health',
'emergency',
'demonstrated',
'our',
'lack',
'of',
'scientific',
'consideration',
'and',
'underlined',
'the',
'alarming',
'need',
'for',
'robust',
'innovations',
'in',
'our',
'health',
'and',
'medical',
'facilities',
'.',
```

```
'For',
'the',
'past',
'few',
'years',
',',
'artificial',
'intelligence',
'has',
'proven',
'to',
'be',
'of',
'tangible',
'potential',
'in',
'the',
'healthcare',
'sectors',
',',
'clinical',
'practices',
',',
'translational',
'medical',
'and',
'biomedical',
'research',
'.',
'After',
'the',
'first',
'case',
'was',
'detected',
'in',
'China',
'on',
'December',
'31st',
'2019',
',',
'it',
'was',
'an',
'AI',
'program',
'developed',
'by',
'BlueDot',
```

```
'that',
'alerted',
'the',
'world',
'about',
'the',
'pandemic',
'.',
'It',
'was',
'quick',
'to',
'realise',
'AI',
'’',
's',
'ability',
'to',
'analyse',
'large',
'chunks',
'of',
'data',
'could',
'help',
'in',
'detecting',
'patterns',
'and',
'identifying',
'and',
'tracking',
'the',
'possible',
'carriers',
'of',
'the',
'virus',
'.',
'Many',
'tracing',
'apps',
'use',
'AI',
'to',
'keep',
'tabs',
'on',
'the',
'people',
```

```
'who',
'have',
'been',
'infected',
'and',
'prevent',
'the',
'risk',
'of',
'cross-infection',
'by',
'using',
'AI',
'algorithms',
'that',
'can',
'track',
'patterns',
'and',
'extract',
'some',
'features',
'to',
'classify',
'or',
'categorise',
'them',
'.',
'So',
'how',
'does',
'AI',
'do',
'that',
'?',
'IBM',
'Watson',
',',
'a',
'sophisticated',
'AI',
'that',
'works',
'on',
'cloud',
'computing',
'and',
'natural',
'language',
'processing',
```

```
',',
'has',
'prominently',
'contributed',
'to',
'the',
'healthcare',
'sector',
'on',
'a',
'global',
'level',
'.',
'Being',
'a',
'conversational',
'AI',
',',
'since',
'2013',
',',
'Watson',
'has',
'helped',
'in',
'recommending',
'treatments',
'to',
'patients',
'suffering',
'from',
'cancer',
'to',
'ensure',
'that',
'they',
'get',
'the',
'best',
'treatment',
'at',
'optimum',
'costs',
'.',
'Researchers',
'at',
'Google',
'Inc.',
'showed',
'that',
```

```
'an',
'AI',
'system',
'can',
'be',
'trained',
'on',
'thousands',
'of',
'images',
'to',
'achieve',
'physician-level',
'sensitivity',
'.',
'By',
'identifying',
'the',
'molecular',
'patterns',
'associated',
'with',
'disease',
'status',
'and',
'its',
'subtypes',
',',
'gene',
'expression',
',',
'and',
'protein',
'abundance',
'levels',
',',
'machine',
'learning',
'methods',
'can',
'detect',
'fatal',
'diseases',
'like',
'cancer',
'at',
'an',
'early',
'stage',
'.',
```

```
'Machine',
'Learning',
'(',
'ML',
')',
'techniques',
'focus',
'mainly',
'on',
'analyzing',
'structured',
'data',
',',
'which',
'can',
'further',
'help',
'in',
'clustering',
'patients',
'’',
'traits',
'and',
'infer',
'the',
'probability',
'of',
'disease',
'outcomes',
'.',
'Since',
'patient',
'traits',
'mainly',
'include',
'masses',
'of',
'data',
'relating',
'to',
'age',
',',
'gender',
',',
'disease',
'history',
',',
'disease-specific',
'data',
'like',
```

```
'diagnostic',
'imaging',
'and',
'gene',
'expressions',
',',
'etc',
',',
'ML',
'can',
'extract',
'features',
'from',
'these',
'data',
'inputs',
'by',
'constructing',
'data',
'analytical',
'algorithms',
'.',
'ML',
'algorithms',
'are',
'either',
'supervised',
'or',
'unsupervised',
'.',
'Unsupervised',
'learning',
'helps',
'in',
'extracting',
'features',
'and',
'clustering',
'similar',
'features',
'together',
'that',
'further',
'leads',
'to',
'early',
'detection',
'of',
'diseases',
'.',
```

```
'Clustering',
'and',
'principal',
'component',
'analysis',
'enable',
'grouping',
'or',
'clustering',
'of',
'similar',
'traits',
'together',
'that',
'are',
'further',
'used',
'to',
'maximize',
'or',
'minimize',
'the',
'similarity',
'between',
'the',
'patients',
'within',
'or',
'between',
'the',
'clusters',
'.',
'Since',
'patient',
'traits',
'are',
'recorded',
'in',
'multiple',
'dimensions',
',',
'such',
'as',
'genes',
',',
'principal',
'component',
'analysis',
'(',
'PCA',
```

```
')',
'creates',
'the',
'apparatus',
'to',
'reduce',
'these',
'dimensions',
'which',
'humans',
'could',
'have',
'not',
'done',
'alone',
'.',
'Supervised',
'learning',
'considers',
'the',
'outcomes',
'of',
'the',
'subjects',
'together',
'with',
'the',
'traits',
',',
'and',
'further',
'correlates',
'the',
'inputs',
'with',
'the',
'outputs',
'to',
'predict',
'the',
'probability',
'of',
'getting',
'a',
'particular',
'clinical',
'event',
',',
'expected',
'value',
```

```
'of',
'a',
'disease',
'level',
'or',
'expected',
'survival',
'time',
',',
'or',
'risk',
'of',
'Down',
'’',
's',
'syndrome',
'.',
'Biomarker',
'panels',
'that',
'are',
'mostly',
'used',
'to',
'detect',
'ovarian',
'cancer',
',',
'have',
'outperformed',
'the',
'conventional',
'statistical',
'methods',
'due',
'to',
'machine',
'learning',
'.',
'In',
'addition',
'to',
'this',
',',
'the',
'use',
'of',
'EHRs',
'and',
'Bayesian',
```

```
'networks',
',',
'which',
'are',
'a',
'part',
'of',
'supervised',
'machine',
'learning',
'algorithms',
',',
'can',
'predict',
'clinical',
'outcomes',
'and',
'mortality',
'respectively',
'.',
'Unstructured',
'data',
'such',
'as',
'clinical',
'notes',
'and',
'texts',
'are',
'converted',
'into',
'machine-readable',
'structured',
'data',
'with',
'the',
'help',
'of',
'natural',
'language',
'processing',
'(',
'NLP',
')',
'.',
'NLP',
'works',
'with',
'two',
'components',
```

```
':',
'text',
'processing',
'and',
'classification',
'.',
'Text',
'processing',
'helps',
'in',
'identifying',
'a',
'series',
'of',
'disease-relevant',
'keywords',
'in',
'clinical',
'notes',
'and',
'then',
'through',
'classification',
'are',
'further',
'categorized',
'into',
'normal',
'and',
'abnormal',
'cases',
'.',
'Chest',
'screening',
'through',
'ML',
'and',
'NLP',
'has',
'helped',
'find',
'abnormalities',
'in',
'the',
'lungs',
'and',
'provide',
'treatment',
'to',
'covid',
```

```
'patients',
'.',
'Healthcare',
'organizations',
'use',
'NLP-based',
'chatbots',
'to',
'increase',
'interactions',
'with',
'patients',
',',
'keeping',
'their',
'mental',
'health',
'and',
'wellness',
'in',
'check',
'.',
'Deep',
'learning',
'is',
'a',
'modern',
'extension',
'of',
'the',
'classical',
'neural',
'network',
'techniques',
'which',
'helps',
'explore',
'more',
'complex',
'non-linear',
'patterns',
'in',
'data',
',',
'using',
'algorithms',
'like',
'convolution',
'neural',
'network',
```

```
',',
'recurrent',
'neural',
'network',
',',
'deep',
'belief',
'network',
',',
'and',
'deep',
'neural',
'network',
'which',
'enables',
'more',
'accurate',
'clinical',
'prediction',
'.',
'When',
'it',
'comes',
'to',
'genome',
'interpretation',
',',
'deep',
'neural',
'networks',
'surpass',
'the',
'conventional',
'methods',
'of',
'logistics',
'regression',
'and',
'support',
'vector',
'machines',
'.',
'Sepsis',
'Watch',
'is',
'an',
'AI',
'system',
'trained',
'in',
```

```
'deep',
'learning',
'algorithms',
'that',
'holds',
'the',
'capability',
'to',
'analyze',
'over',
'32',
'million',
'data',
'points',
'to',
'create',
'a',
'patient',
'’',
's',
'risk',
'score',
'and',
'identify',
'the',
'early',
'stages',
'of',
'sepsis',
'.',
'Another',
'method',
'known',
'as',
'the',
'Learning-based',
'Optimization',
'of',
'the',
'Under',
'Sampling',
'Pattern',
'(',
'LOUPE',
')',
'is',
'based',
'on',
'integrating',
'full',
```

```
'resolution',
'MRI',
'scans',
'with',
'the',
'convolutional',
'neural',
'network',
'algorithm',
',',
'which',
'helps',
'in',
'creating',
'more',
'accurate',
'reconstructions',
'.',
'Robotic',
'surgery',
'is',
'widely',
'considered',
'in',
'most',
'delicate',
'surgeries',
'like',
'gynaecology',
'and',
'prostate',
'surgery',
'.',
'Even',
'after',
'striking',
'the',
'right',
'balance',
'between',
'human',
'decisions',
'and',
'AI',
'precision',
',',
'robotic',
'surgery',
'reduces',
'surgeon',
```

```
'efficiency',
'as',
'they',
'have',
'to',
'be',
'manually',
'operated',
'through',
'a',
'console',
'.',
'Thus',
',',
'autonomous',
'robotic',
'surgery',
'is',
'on',
'the',
'rise',
'with',
'inventions',
'such',
'as',
'robotic',
'silicon',
'fingers',
'that',
'mimic',
'the',
'sense',
'of',
'touch',
'that',
'surgeons',
'need',
'to',
'identify',
'organs',
',',
'cut',
'tissues',
',',
'etc.',
',',
'or',
'robotic',
'catheters',
'that',
```

```
'can',
'navigate',
'whether',
'it',
'is',
'touching',
'blood',
',',
'tissue',
',',
'or',
'valve',
'.',
'Researchers',
'at',
'Children',
'’',
's',
'National',
'Hospital',
',',
'Washington',
'have',
'already',
'developed',
'an',
'AI',
'called',
'Smart',
'Tissue',
'Autonomous',
'Robot',
'(',
'STAR',
')',
',',
'which',
'performs',
'a',
'colon',
'anastomosis',
'on',
'its',
'own',
'with',
'the',
'help',
'of',
...]
```

```
tokenize_words_without_stopwords = []

for word in tokens:
  if word not in stop_words:
    tokenize_words_without_stopwords.append(word)
print(tokenize_words_without_stopwords)
```

```
['Introduction', '“', 'If', 'anything', 'kills', '10', 'million',
'people', 'next', 'decades', ',', 'highly', 'infectious', 'virus',
'rather', 'war', '.', 'Not', 'missiles', 'microbes.', '”', 'Bill',
'Gates', '’', 'remarks', 'TED', 'conference', '2014', ',', 'right',
'world', 'avoided', 'Ebola', 'outbreak', '.', 'When', 'new', ',',
'unprecedented', ',', 'invisible', 'virus', 'hit', 'us', ',', 'met',
'overwhelmed', 'unprepared', 'healthcare', 'system', 'oblivious',
'population', '.', 'This', 'public', 'health', 'emergency',
'demonstrated', 'lack', 'scientific', 'consideration', 'underlined',
'alarming', 'need', 'robust', 'innovations', 'health', 'medical',
'facilities', '.', 'For', 'past', 'years', ',', 'artificial',
'intelligence', 'proven', 'tangible', 'potential', 'healthcare',
'sectors', ',', 'clinical', 'practices', ',', 'translational',
'medical', 'biomedical', 'research', '.', 'After', 'first', 'case',
'detected', 'China', 'December', '31st', '2019', ',', 'AI', 'program',
'developed', 'BlueDot', 'alerted', 'world', 'pandemic', '.', 'It',
'quick', 'realise', 'AI', '’', 'ability', 'analyse', 'large',
'chunks', 'data', 'could', 'help', 'detecting', 'patterns',
'identifying', 'tracking', 'possible', 'carriers', 'virus', '.',
'Many', 'tracing', 'apps', 'use', 'AI', 'keep', 'tabs', 'people',
'infected', 'prevent', 'risk', 'cross-infection', 'using', 'AI',
'algorithms', 'track', 'patterns', 'extract', 'features', 'classify',
'categorise', '.', 'So', 'AI', '?', 'IBM', 'Watson', ',',
'sophisticated', 'AI', 'works', 'cloud', 'computing', 'natural',
'language', 'processing', ',', 'prominently', 'contributed',
'healthcare', 'sector', 'global', 'level', '.', 'Being',
'conversational', 'AI', ',', 'since', '2013', ',', 'Watson', 'helped',
'recommending', 'treatments', 'patients', 'suffering', 'cancer',
'ensure', 'get', 'best', 'treatment', 'optimum', 'costs', '.',
'Researchers', 'Google', 'Inc.', 'showed', 'AI', 'system', 'trained',
'thousands', 'images', 'achieve', 'physician-level', 'sensitivity',
'.', 'By', 'identifying', 'molecular', 'patterns', 'associated',
'disease', 'status', 'subtypes', ',', 'gene', 'expression', ',',
'protein', 'abundance', 'levels', ',', 'machine', 'learning',
'methods', 'detect', 'fatal', 'diseases', 'like', 'cancer', 'early',
'stage', '.', 'Machine', 'Learning', '(', 'ML', ')', 'techniques',
'focus', 'mainly', 'analyzing', 'structured', 'data', ',', 'help',
'clustering', 'patients', '’', 'traits', 'infer', 'probability',
'disease', 'outcomes', '.', 'Since', 'patient', 'traits', 'mainly',
'include', 'masses', 'data', 'relating', 'age', ',', 'gender', ',',
'disease', 'history', ',', 'disease-specific', 'data', 'like',
'diagnostic', 'imaging', 'gene', 'expressions', ',', 'etc', ',', 'ML',
'extract', 'features', 'data', 'inputs', 'constructing', 'data',
```

'analytical', 'algorithms', '.', 'ML', 'algorithms', 'either',
'supervised', 'unsupervised', '.', 'Unsupervised', 'learning',
'helps', 'extracting', 'features', 'clustering', 'similar',
'features', 'together', 'leads', 'early', 'detection', 'diseases',
'.', 'Clustering', 'principal', 'component', 'analysis', 'enable',
'grouping', 'clustering', 'similar', 'traits', 'together', 'used',
'maximize', 'minimize', 'similarity', 'patients', 'within',
'clusters', '.', 'Since', 'patient', 'traits', 'recorded', 'multiple',
'dimensions', ',', 'genes', ',', 'principal', 'component', 'analysis',
'(', 'PCA', ')', 'creates', 'apparatus', 'reduce', 'dimensions',
'humans', 'could', 'done', 'alone', '.', 'Supervised', 'learning',
'considers', 'outcomes', 'subjects', 'together', 'traits', ',',
'correlates', 'inputs', 'outputs', 'predict', 'probability',
'getting', 'particular', 'clinical', 'event', ',', 'expected',
'value', 'disease', 'level', 'expected', 'survival', 'time', ',',
'risk', 'Down', '’', 'syndrome', '.', 'Biomarker', 'panels', 'mostly',
'used', 'detect', 'ovarian', 'cancer', ',', 'outperformed',
'conventional', 'statistical', 'methods', 'due', 'machine',
'learning', '.', 'In', 'addition', ',', 'use', 'EHRs', 'Bayesian',
'networks', ',', 'part', 'supervised', 'machine', 'learning',
'algorithms', ',', 'predict', 'clinical', 'outcomes', 'mortality',
'respectively', '.', 'Unstructured', 'data', 'clinical', 'notes',
'texts', 'converted', 'machine-readable', 'structured', 'data',
'help', 'natural', 'language', 'processing', '(', 'NLP', ')', '.',
'NLP', 'works', 'two', 'components', ':', 'text', 'processing',
'classification', '.', 'Text', 'processing', 'helps', 'identifying',
'series', 'disease-relevant', 'keywords', 'clinical', 'notes',
'classification', 'categorized', 'normal', 'abnormal', 'cases', '.',
'Chest', 'screening', 'ML', 'NLP', 'helped', 'find', 'abnormalities',
'lungs', 'provide', 'treatment', 'covid', 'patients', '.',
'Healthcare', 'organizations', 'use', 'NLP-based', 'chatbots',
'increase', 'interactions', 'patients', ',', 'keeping', 'mental',
'health', 'wellness', 'check', '.', 'Deep', 'learning', 'modern',
'extension', 'classical', 'neural', 'network', 'techniques', 'helps',
'explore', 'complex', 'non-linear', 'patterns', 'data', ',', 'using',
'algorithms', 'like', 'convolution', 'neural', 'network', ',',
'recurrent', 'neural', 'network', ',', 'deep', 'belief', 'network',
',', 'deep', 'neural', 'network', 'enables', 'accurate', 'clinical',
'prediction', '.', 'When', 'comes', 'genome', 'interpretation', ',',
'deep', 'neural', 'networks', 'surpass', 'conventional', 'methods',
'logistics', 'regression', 'support', 'vector', 'machines', '.',
'Sepsis', 'Watch', 'AI', 'system', 'trained', 'deep', 'learning',
'algorithms', 'holds', 'capability', 'analyze', '32', 'million',
'data', 'points', 'create', 'patient', '’', 'risk', 'score',
'identify', 'early', 'stages', 'sepsis', '.', 'Another', 'method',
'known', 'Learning-based', 'Optimization', 'Under', 'Sampling',
'Pattern', '(', 'LOUPE', ')', 'based', 'integrating', 'full',
'resolution', 'MRI', 'scans', 'convolutional', 'neural', 'network',
'algorithm', ',', 'helps', 'creating', 'accurate', 'reconstructions',
'.', 'Robotic', 'surgery', 'widely', 'considered', 'delicate',

'surgeries', 'like', 'gynaecology', 'prostate', 'surgery', '.',
'Even', 'striking', 'right', 'balance', 'human', 'decisions', 'AI',
'precision', ',', 'robotic', 'surgery', 'reduces', 'surgeon',
'efficiency', 'manually', 'operated', 'console', '.', 'Thus', ',',
'autonomous', 'robotic', 'surgery', 'rise', 'inventions', 'robotic',
'silicon', 'fingers', 'mimic', 'sense', 'touch', 'surgeons', 'need',
'identify', 'organs', ',', 'cut', 'tissues', ',', 'etc.', ',',
'robotic', 'catheters', 'navigate', 'whether', 'touching', 'blood',
',', 'tissue', ',', 'valve', '.', 'Researchers', 'Children', '’',
'National', 'Hospital', ',', 'Washington', 'already', 'developed',
'AI', 'called', 'Smart', 'Tissue', 'Autonomous', 'Robot', '(', 'STAR',
')', ',', 'performs', 'colon', 'anastomosis', 'help', 'ML-powered',
'suturing', 'tool', ',', 'automatically', 'detects', 'patient', '’',
'breathing', 'pattern', 'apply', 'suture', 'correct', 'point', '.',
'An', 'image', 'STAR', 'surgery', '.', 'Cloud', 'computing',
'healthcare', 'helped', 'retrieving', 'sharing', 'medical', 'records',
'safely', 'reduction', 'maintenance', 'costs', '.', 'Through',
'technology', 'doctors', 'various', 'healthcare', 'workers', 'access',
'detailed', 'patient', 'data', 'helps', 'speeding', 'analysis',
'ultimately', 'leading', 'better', 'care', 'form', 'accurate',
'information', ',', 'medications', ',', 'therapies', '.', 'How', 'It',
'help', 'Biomedical', 'research', '?', 'Since', 'AI', 'analyze',
'literature', 'beyond', 'readability', ',', 'used', 'concise',
'biomedical', 'research', '.', 'With', 'help', 'ML', 'algorithms',
'NLP', ',', 'AI', 'accelerate', 'screening', 'indexing', 'biomedical',
'research', ',', 'ranking', 'literature', 'interest', 'allows',
'researchers', 'formulate', 'test', 'scientific', 'hypotheses', 'far',
'precisely', 'quickly', '.', 'Taking', 'next', 'level', ',', 'AI',
'systems', 'like', 'computational', 'modelling', 'assistant', '(',
'CMA', ')', 'helps', 'researchers', 'construct', 'simulation',
'models', 'concepts', 'mind', '.', 'Such', 'innovations', 'majorly',
'contributed', 'topics', 'tumour', 'suppressor', 'mechanisms',
'protein-protein', 'interaction', 'information', 'extraction', '.',
'AI', 'precision', 'medicine', 'Since', 'precision', 'medicine',
'focuses', 'healthcare', 'interventions', 'individuals', 'groups',
'patients', 'based', 'profile', ',', 'various', 'AI', 'devices',
'pave', 'way', 'practice', 'efficiently', '.', 'With', 'help', 'ML',
',', 'complex', 'algorithms', 'like', 'large', 'datasets', 'used',
'predict', 'create', 'optimal', 'treatment', 'strategy', '.', 'Deep',
'learning', 'neural', 'networks', 'used', 'process', 'data',
'healthcare', 'apps', 'keep', 'close', 'watch', 'patient', '’',
'emotional', 'state', ',', 'food', 'intake', ',', 'health',
'monitoring', '.', '“', 'Omics', '”', 'refers', 'collective',
'technologies', 'help', 'exploring', 'roles', ',', 'relationships',
'various', 'branches', 'ending', 'suffix', '“', 'omics', '”',
'genomics', ',', 'proteomics', ',', 'etc', '.', 'Omics-based',
'tests', 'based', 'machine', 'learning', 'algorithms', 'help', 'find',
'correlations', 'predict', 'treatment', 'responses', ',',
'ultimately', 'creating', 'personalized', 'treatments', 'individual',
'patients', '.', 'How', 'helps', 'psychology', 'neuro', 'patients',

'For', 'psychologists', 'studying', 'creativity', ',', 'AI',
'promising', 'new', 'classes', 'experiments', 'developing', 'data',
'structures', 'programs', 'exploring', 'novel', 'theories', 'new',
'horizon', '.', 'Studies', 'show', 'AI', 'conduct', 'therapy',
'sessions', ',', 'e-therapy', 'sessions', ',', 'assessments',
'autonomously', ',', 'also', 'assisting', 'human', 'practitioners',
',', ',', 'sessions', '.', 'The', 'Detection', 'Computational',
'Analysis', 'Psychological', 'Signal', 'project', 'uses', 'ML', ',',
'computer', 'vision', ',', 'NLP', 'analyze', 'language', ',',
'physical', 'gestures', ',', 'social', 'signals', 'identify', 'cues',
'human', 'distress', '.', 'This', 'ground-breaking', 'technology',
'assesses', 'soldiers', 'returning', 'combat', 'recognizes',
'require', 'mental', 'health', 'support', '.', 'In', 'future', ',',
'combine', 'data', 'captured', 'face-to-face', 'interviews',
'information', 'sleeping', ',', 'eating', ',', 'online', 'behaviours',
'complete', 'patient', 'view', '.', 'Stroke', 'identification',
'Stroke', 'another', 'frequently', 'occurring', 'disease', 'affects',
'500', 'million', 'people', 'worldwide', '.', 'Thrombus', ',',
'vessel', 'cerebral', 'infarction', 'major', '(', '85', '%', ')',
'cause', 'stroke', 'occurrence', '.', 'In', 'recent', 'years', ',',
'AI', 'techniques', 'used', 'numerous', 'stroke-related', 'studies',
'early', 'detection', 'timely', 'treatment', 'along', 'efficient',
'outcome', 'prediction', 'help', 'solve', 'problem', '.', 'With',
'AI', 'disposal', ',', 'large', 'amounts', 'data', 'rich',
'information', ',', 'complications', 'real-life', 'clinical',
'questions', 'addressed', 'arena', '.', 'Currently', ',', 'two', 'ML',
'algorithms-', 'genetic', 'fuzzy', 'finite', 'state', 'machine',
'PCA', 'implemented', 'build', 'model', 'building', 'solution', '.',
'These', 'include', 'human', 'activity', 'recognition', 'stage',
'stroke', 'onset', 'detection', 'stage', '.', 'An', 'alert', 'stroke',
'message', 'activated', 'soon', 'movement', 'significantly',
'different', 'normal', 'pattern', 'recorded', '.', 'ML', 'methods',
'applied', 'neuroimaging', 'data', 'assist', 'disease', 'evaluation',
'predicting', 'stroke', 'treatment', 'diagnosis', '.', 'Patient',
'Monitoring', 'Today', ',', 'market', 'AI-based', 'patient',
'monitoring', 'impressive', 'monetarily', 'enticing', '.', 'It',
'evolving', 'artificial', 'sensors', ',', 'smart', 'technologies',
'explores', 'everything', 'brain-computer', 'interfaces',
'nanorobotics', '.', 'Companies', 'smart-watches', 'engaged',
'people', 'perform', 'remote', 'monitoring', 'even', '“', 'patients',
'”', '.', 'An', 'obvious', 'place', 'start', 'wearable', 'embedded',
'sensors', ',', 'glucose', 'monitors', ',', 'pulse', 'monitors', ',',
'oximeters', ',', 'ECG', 'monitors', '.', 'With', 'patient',
'monitoring', 'becoming', 'crucial', ',', 'AI', 'finds', 'numerous',
'applications', 'chronic', 'conditions', ',', 'intensive', 'care',
'units', ',', 'operating', 'rooms', ',', 'emergency', 'rooms', ',',
'cardiac', 'wards', 'timeless', 'clinical', 'decision-making',
'measured', 'seconds', '.', 'More', 'advances', 'started', 'gain',
'traction', 'like', 'smart', 'prosthetics', 'implants', '.', 'These',
'play', 'impeccable', 'role', 'patient', 'management', 'post-surgery',

```
'rehabilitation', '.', 'Demographics', ',', 'laboratory', 'results',
'vital', 'signs', 'also', 'used', 'predict', 'cardiac', 'arrest', ',',
'transfer', 'intensive', 'care', 'unit', ',', 'even', 'death', '.',
'In', 'addition', ',', 'interpretable', 'machine-learning', 'model',
'assist', 'anesthesiologists', 'predicting', 'hypoxaemia', 'events',
'surgery', '.', 'This', 'suggests', 'deep-learning', 'algorithms',
',', 'raw', 'patient-monitoring', 'data', 'could', 'better', 'used',
'avoid', 'information', 'overload', 'alert', 'overload', 'enabling',
'accurate', 'clinical', 'prediction', 'timely', 'decision-making',
'.', 'Conclusion', 'Considering', 'vast', 'range', 'tasks', 'AI', ',',
'evident', 'holds', 'deep', 'potential', 'improving', 'patient',
'outcomes', 'skyrocketing', 'levels', '.', 'Using', 'sophisticated',
'algorithms', 'AI', 'bring', 'revolution', 'healthcare', 'sector',
'.', 'Even', 'facing', 'challenges', 'like', 'whether', 'technology',
'able', 'deliver', 'promises', ',', 'ethical', 'measures', ',',
'training', 'physicians', 'use', ',', 'standard', 'regulations',
'etc', ',', 'role', 'AI', 'transforming', 'clinical', 'practices',
'ignored', '.', 'The', 'biggest', 'challenge', 'integration', 'AI',
'daily', 'practice', '.', 'All', 'overcome', 'within', 'period',
'technologies', 'mature', 'making', 'system', 'far', 'enhanced',
'effective', '.', 'Blackcoffer', 'Insights', '29', ':', 'Sanskriti',
'Sunderum', 'Aayushi', 'Nauhwar', ',', 'SRCC', ',', 'Delhi',
'University']
```

```
def remove_stopwords():
  for sentense in list(df['Context']):
    tokens = word_tokenize(sentense)
    tokenize_words_without_stopwords = []
    for word in tokens:
      if word not in stop_words:
        tokenize_words_without_stopwords.append(word)
    return tokenize_words_without_stopwords
```

###Calculating word count after removing stop words

```
count_word = []
for i in df['Context']:
  tokens = word_tokenize(i)
  tokenize_words_without_stopwords = []
  for word in tokens:
    if word not in stop_words:
      tokenize_words_without_stopwords.append(word)
  count_word.append(len(tokenize_words_without_stopwords))

df["Count of words"] = count_word
```

###Creating Ids

```
df = df.reset_index()
df = df.rename(columns={'index':'Ids'})
df
```

```
     Ids                                              Websites  \
0      0  https://insights.blackcoffer.com/ai-in-healthc...
1      1  https://insights.blackcoffer.com/what-if-the-c...
2      2  https://insights.blackcoffer.com/what-jobs-wil...
3      3  https://insights.blackcoffer.com/will-machine-...
4      4  https://insights.blackcoffer.com/will-ai-repla...
..   ...                                                ...
109  109  https://insights.blackcoffer.com/blockchain-fo...
110  110  https://insights.blackcoffer.com/the-future-of...
111  111  https://insights.blackcoffer.com/big-data-anal...
112  112  https://insights.blackcoffer.com/business-anal...
113  113  https://insights.blackcoffer.com/challenges-an...

                                               Context  Count of words

0     Introduction “If anything kills over 10 milli...            1396

1     Human minds, a fascination in itself carrying...            1012

2     Introduction AI is rapidly evolving in the em...            1255

3     “Anything that could give rise to smarter-tha...            1157

4     “Machine intelligence is the last invention t...            1253

..                                                ...             ...

109   Reconciling with the financial realities of a...             672

110   What Is an Investment? An investment is a res...            1212

111   Quality and affordable healthcare is a vision...             847

112   Analytics is a statistical scientific process...             516

113   Big Data To begin with I shall first like to ...             733


[114 rows x 4 columns]
```

###Creating sentiment analyzer

```
from nltk.sentiment import SentimentIntensityAnalyzer
from tqdm.notebook import tqdm
nltk.download('vader_lexicon')

sia = SentimentIntensityAnalyzer()
```

```
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
```

```
sia.polarity_scores(example) ##Checking example
```

```
{'neg': 0.044, 'neu': 0.811, 'pos': 0.145, 'compound': 0.9996}
```

###Applying the analyzer on entire dataset

```
# Run the polarity score on the entire dataset
res = {}
for i, row in tqdm(df.iterrows(), total=len(df)):
    text = row['Context']
    myid = row['Ids']
    res[myid] = sia.polarity_scores(text)
```


```
{"model_id":"92114983ed7d4e2b81a9b39120854af8","version_major":2,"vers
ion_minor":0}
```


###Merging the result with dataset

```
vaders = pd.DataFrame(res).T
vaders = vaders.reset_index().rename(columns={'index': 'Ids'})
vaders = vaders.merge(df, how='left')
```

###Renaming the dataset

```
df = vaders.iloc[:,:]
df
```

```
     Ids    neg    neu    pos  compound  \
0      0  0.044  0.811  0.145    0.9996
1      1  0.084  0.767  0.149    0.9984
2      2  0.036  0.889  0.075    0.9891
3      3  0.044  0.811  0.145    0.9995
4      4  0.050  0.804  0.147    0.9996
..   ...    ...    ...    ...       ...
109  109  0.051  0.820  0.129    0.9973
110  110  0.020  0.871  0.110    0.9993
111  111  0.092  0.818  0.091   -0.9301
112  112  0.004  0.825  0.170    0.9989
113  113  0.082  0.732  0.186    0.9990

                                              Websites  \
0    https://insights.blackcoffer.com/ai-in-healthc...
1    https://insights.blackcoffer.com/what-if-the-c...
2    https://insights.blackcoffer.com/what-jobs-wil...
3    https://insights.blackcoffer.com/will-machine-...
4    https://insights.blackcoffer.com/will-ai-repla...
..                                                 ...
109  https://insights.blackcoffer.com/blockchain-fo...
110  https://insights.blackcoffer.com/the-future-of...
111  https://insights.blackcoffer.com/big-data-anal...
112  https://insights.blackcoffer.com/business-anal...
113  https://insights.blackcoffer.com/challenges-an...
```

```
                                                  Context  Count of words
0     Introduction “If anything kills over 10 milli...            1396
1     Human minds, a fascination in itself carrying...            1012
2     Introduction AI is rapidly evolving in the em...            1255
3     “Anything that could give rise to smarter-tha...            1157
4     “Machine intelligence is the last invention t...            1253
..                                                  ...             ...
109   Reconciling with the financial realities of a...             672
110   What Is an Investment? An investment is a res...            1212
111   Quality and affordable healthcare is a vision...             847
112   Analytics is a statistical scientific process...             516
113   Big Data To begin with I shall first like to ...             733

[114 rows x 8 columns]
```

##Polarity Score

```
polarity_score = []
for i, row in tqdm(df.iterrows(), total=len(df)):
    pos = row['pos']
    neg = row['neg']
    polarity = (pos - neg)/ ((pos + neg) + 0.000001)
    polarity_score.append(polarity)
```

{"model_id":"6ded1e12e6c84e7ba86f8ee83aac6b80","version_major":2,"version_minor":0}

```
df["Polarity Score"] = polarity_score
df
```

```
     Ids    neg    neu    pos  compound  \
0      0  0.044  0.811  0.145    0.9996
1      1  0.084  0.767  0.149    0.9984
2      2  0.036  0.889  0.075    0.9891
3      3  0.044  0.811  0.145    0.9995
4      4  0.050  0.804  0.147    0.9996
..   ...    ...    ...    ...       ...
```

109	109	0.051	0.820	0.129	0.9973
110	110	0.020	0.871	0.110	0.9993
111	111	0.092	0.818	0.091	-0.9301
112	112	0.004	0.825	0.170	0.9989
113	113	0.082	0.732	0.186	0.9990

	Websites
0	https://insights.blackcoffer.com/ai-in-healthc...
1	https://insights.blackcoffer.com/what-if-the-c...
2	https://insights.blackcoffer.com/what-jobs-wil...
3	https://insights.blackcoffer.com/will-machine-...
4	https://insights.blackcoffer.com/will-ai-repla...
..	...
109	https://insights.blackcoffer.com/blockchain-fo...
110	https://insights.blackcoffer.com/the-future-of...
111	https://insights.blackcoffer.com/big-data-anal...
112	https://insights.blackcoffer.com/business-anal...
113	https://insights.blackcoffer.com/challenges-an...

	Context	Count of words
0	Introduction “If anything kills over 10 milli...	1396
1	Human minds, a fascination in itself carrying...	1012
2	Introduction AI is rapidly evolving in the em...	1255
3	“Anything that could give rise to smarter-tha...	1157
4	“Machine intelligence is the last invention t...	1253
..	...	...
109	Reconciling with the financial realities of a...	672
110	What Is an Investment? An investment is a res...	1212
111	Quality and affordable healthcare is a vision...	847
112	Analytics is a statistical scientific process...	516
113	Big Data To begin with I shall first like to ...	733

	Polarity Score
0	0.534389
1	0.278969
2	0.351348
3	0.534389

```
4          0.492383
..              ...
109        0.433331
110        0.692302
111       -0.005464
112        0.954018
113        0.388058

[114 rows x 9 columns]
```

##Subjectivity Score

```
subjectivity_score = []
for i, row in tqdm(df.iterrows(), total=len(df)):
    pos = row['pos']
    neg = row['neg']
    total_word = row['Count of words']
    sub = (pos + neg)/ ((total_word) + 0.000001)
    subjectivity_score.append(sub)
```

{"model_id":"b1a2d890d5b9479b9f79fc74fe0d4768","version_major":2,"vers
ion_minor":0}

```
df["Subjectivity Score"] = subjectivity_score
df
```

```
     Ids    neg    neu    pos  compound  \
0      0  0.044  0.811  0.145    0.9996
1      1  0.084  0.767  0.149    0.9984
2      2  0.036  0.889  0.075    0.9891
3      3  0.044  0.811  0.145    0.9995
4      4  0.050  0.804  0.147    0.9996
..   ...    ...    ...    ...       ...
109  109  0.051  0.820  0.129    0.9973
110  110  0.020  0.871  0.110    0.9993
111  111  0.092  0.818  0.091   -0.9301
112  112  0.004  0.825  0.170    0.9989
113  113  0.082  0.732  0.186    0.9990

                                              Websites  \
0    https://insights.blackcoffer.com/ai-in-healthc...
1    https://insights.blackcoffer.com/what-if-the-c...
2    https://insights.blackcoffer.com/what-jobs-wil...
3    https://insights.blackcoffer.com/will-machine-...
4    https://insights.blackcoffer.com/will-ai-repla...
..                                                 ...
109  https://insights.blackcoffer.com/blockchain-fo...
110  https://insights.blackcoffer.com/the-future-of...
111  https://insights.blackcoffer.com/big-data-anal...
112  https://insights.blackcoffer.com/business-anal...
113  https://insights.blackcoffer.com/challenges-an...
```

```
                                                 Context  Count of words
\
0     Introduction “If anything kills over 10 milli...            1396

1     Human minds, a fascination in itself carrying...            1012

2     Introduction AI is rapidly evolving in the em...            1255

3     “Anything that could give rise to smarter-tha...            1157

4     “Machine intelligence is the last invention t...            1253

..                                                ...             ...

109   Reconciling with the financial realities of a...             672

110   What Is an Investment? An investment is a res...            1212

111   Quality and affordable healthcare is a vision...             847

112   Analytics is a statistical scientific process...             516

113   Big Data To begin with I shall first like to ...             733


     Polarity Score  Subjectivity Score
0          0.534389            0.000135
1          0.278969            0.000230
2          0.351348            0.000088
3          0.534389            0.000163
4          0.492383            0.000157
..              ...                 ...
109        0.433331            0.000268
110        0.692302            0.000107
111       -0.005464            0.000216
112        0.954018            0.000337
113        0.388058            0.000366

[114 rows x 10 columns]
```

###Importing libraries for sentense count

```
from nltk.tokenize import sent_tokenize

texts = df['Context']
count_sentense = []
for text in texts:
```

```
  total_sentense = sent_tokenize(text)
  count_sentense.append(len(total_sentense))

df['number_sentense'] = count_sentense
```

##Average Sentense Length

```
average_sentense_length = []
for i, row in tqdm(df.iterrows(), total=len(df)):
    number_words = row['Count of words']
    number_sentense = row['number_sentense']
    sentense_length = round((number_words / number_sentense),2)
    average_sentense_length.append(sentense_length)
```

{"model_id":"9c1f210897b2418db3e0b16526cc94fd","version_major":2,"vers
ion_minor":0}

```
df['Average Sentense Length'] = average_sentense_length
df
```

```
     Ids    neg    neu    pos  compound  \
0      0  0.044  0.811  0.145    0.9996
1      1  0.084  0.767  0.149    0.9984
2      2  0.036  0.889  0.075    0.9891
3      3  0.044  0.811  0.145    0.9995
4      4  0.050  0.804  0.147    0.9996
..   ...    ...    ...    ...       ...
109  109  0.051  0.820  0.129    0.9973
110  110  0.020  0.871  0.110    0.9993
111  111  0.092  0.818  0.091   -0.9301
112  112  0.004  0.825  0.170    0.9989
113  113  0.082  0.732  0.186    0.9990

                                              Websites  \
0    https://insights.blackcoffer.com/ai-in-healthc...
1    https://insights.blackcoffer.com/what-if-the-c...
2    https://insights.blackcoffer.com/what-jobs-wil...
3    https://insights.blackcoffer.com/will-machine-...
4    https://insights.blackcoffer.com/will-ai-repla...
..                                                 ...
109  https://insights.blackcoffer.com/blockchain-fo...
110  https://insights.blackcoffer.com/the-future-of...
111  https://insights.blackcoffer.com/big-data-anal...
112  https://insights.blackcoffer.com/business-anal...
113  https://insights.blackcoffer.com/challenges-an...

                                               Context  Count of words
\
0     Introduction “If anything kills over 10 milli...            1396

1     Human minds, a fascination in itself carrying...            1012
```

```
2     Introduction AI is rapidly evolving in the em...           1255

3     “Anything that could give rise to smarter-tha...           1157

4     “Machine intelligence is the last invention t...           1253

..                                                 ...            ...

109   Reconciling with the financial realities of a...            672

110   What Is an Investment? An investment is a res...           1212

111   Quality and affordable healthcare is a vision...            847

112   Analytics is a statistical scientific process...            516

113   Big Data To begin with I shall first like to ...            733


     Polarity Score  Subjectivity Score  number_sentense  \
0          0.534389            0.000135               77
1          0.278969            0.000230               80
2          0.351348            0.000088               85
3          0.534389            0.000163               95
4          0.492383            0.000157               79
..              ...                 ...              ...
109        0.433331            0.000268               49
110        0.692302            0.000107               62
111       -0.005464            0.000216               66
112        0.954018            0.000337               29
113        0.388058            0.000366               66

     Average Sentense Length
0                      18.13
1                      12.65
2                      14.76
3                      12.18
4                      15.86
..                       ...
109                    13.71
110                    19.55
111                    12.83
112                    17.79
113                    11.11

[114 rows x 12 columns]
```

###Importing libraries for syllable word count

```
import spacy
from textstat.textstat import textstatistics
```

##Complex word count

```
complex_word_count = []

for i, row in tqdm(df.iterrows(), total=len(df)):
  words = []
  text = row['Context']
  tokens = word_tokenize(text)

  for word in tokens:
    if word not in stop_words:
      words.append(word)

  diff_words_set = []
  nlp = spacy.load('en_core_web_sm')

  for word in words:
    syllable_count = textstatistics().syllable_count(word)
    if word not in nlp.Defaults.stop_words and syllable_count >= 2:
      diff_words_set.append(word)
  complex_word_count.append(len(diff_words_set))
```

{"model_id":"7d6b3d7be295497d98a11c2cc6ffce33","version_major":2,"vers
ion_minor":0}

```
df['Complex Word Count'] = complex_word_count
df
```

```
     Ids    neg    neu    pos  compound  \
0      0  0.044  0.811  0.145    0.9996
1      1  0.084  0.767  0.149    0.9984
2      2  0.036  0.889  0.075    0.9891
3      3  0.044  0.811  0.145    0.9995
4      4  0.050  0.804  0.147    0.9996
..   ...    ...    ...    ...       ...
109  109  0.051  0.820  0.129    0.9973
110  110  0.020  0.871  0.110    0.9993
111  111  0.092  0.818  0.091   -0.9301
112  112  0.004  0.825  0.170    0.9989
113  113  0.082  0.732  0.186    0.9990

                                              Websites  \
0    https://insights.blackcoffer.com/ai-in-healthc...
1    https://insights.blackcoffer.com/what-if-the-c...
2    https://insights.blackcoffer.com/what-jobs-wil...
3    https://insights.blackcoffer.com/will-machine-...
4    https://insights.blackcoffer.com/will-ai-repla...
..                                                 ...
```

```
109  https://insights.blackcoffer.com/blockchain-fo...
110  https://insights.blackcoffer.com/the-future-of...
111  https://insights.blackcoffer.com/big-data-anal...
112  https://insights.blackcoffer.com/business-anal...
113  https://insights.blackcoffer.com/challenges-an...

                                                Context  Count of words
\
0     Introduction “If anything kills over 10 milli...            1396

1     Human minds, a fascination in itself carrying...            1012

2     Introduction AI is rapidly evolving in the em...            1255

3     “Anything that could give rise to smarter-tha...            1157

4     “Machine intelligence is the last invention t...            1253

..                                                ...             ...

109   Reconciling with the financial realities of a...             672

110   What Is an Investment? An investment is a res...            1212

111   Quality and affordable healthcare is a vision...             847

112   Analytics is a statistical scientific process...             516

113   Big Data To begin with I shall first like to ...             733


     Polarity Score  Subjectivity Score  number_sentense  \
0          0.534389            0.000135               77
1          0.278969            0.000230               80
2          0.351348            0.000088               85
3          0.534389            0.000163               95
4          0.492383            0.000157               79
..              ...                 ...              ...
109        0.433331            0.000268               49
110        0.692302            0.000107               62
111       -0.005464            0.000216               66
112        0.954018            0.000337               29
113        0.388058            0.000366               66

     Average Sentense Length  Complex Word Count
0                      18.13                 798
1                      12.65                 397
2                      14.76                 707
3                      12.18                 513
```

```
4                      15.86                 627
..                       ...                 ...
109                    13.71                 331
110                    19.55                 559
111                    12.83                 460
112                    17.79                 308
113                    11.11                 389

[114 rows x 13 columns]
```

##Percentage of complex words

```
percentage_complex_word = []
for i, row in tqdm(df.iterrows(), total=len(df)):
    number_words = row['Count of words']
    complex_word = row['Complex Word Count']
    percentage = round((complex_word / number_words) * 100,2)
    percentage_complex_word.append(percentage)
```

{"model_id":"12577b6be06e49799140faba6d676d4c","version_major":2,"version_minor":0}

```
df["Percentage of Complex words"] = percentage_complex_word
df
```

```
     Ids    neg    neu    pos  compound  \
0      0  0.044  0.811  0.145    0.9996
1      1  0.084  0.767  0.149    0.9984
2      2  0.036  0.889  0.075    0.9891
3      3  0.044  0.811  0.145    0.9995
4      4  0.050  0.804  0.147    0.9996
..   ...    ...    ...    ...       ...
109  109  0.051  0.820  0.129    0.9973
110  110  0.020  0.871  0.110    0.9993
111  111  0.092  0.818  0.091   -0.9301
112  112  0.004  0.825  0.170    0.9989
113  113  0.082  0.732  0.186    0.9990

                                              Websites  \
0    https://insights.blackcoffer.com/ai-in-healthc...
1    https://insights.blackcoffer.com/what-if-the-c...
2    https://insights.blackcoffer.com/what-jobs-wil...
3    https://insights.blackcoffer.com/will-machine-...
4    https://insights.blackcoffer.com/will-ai-repla...
..                                                 ...
109  https://insights.blackcoffer.com/blockchain-fo...
110  https://insights.blackcoffer.com/the-future-of...
111  https://insights.blackcoffer.com/big-data-anal...
112  https://insights.blackcoffer.com/business-anal...
113  https://insights.blackcoffer.com/challenges-an...
```

	Context	Count of words
0	Introduction “If anything kills over 10 milli...	1396
1	Human minds, a fascination in itself carrying...	1012
2	Introduction AI is rapidly evolving in the em...	1255
3	“Anything that could give rise to smarter-tha...	1157
4	“Machine intelligence is the last invention t...	1253
..	...	...
109	Reconciling with the financial realities of a...	672
110	What Is an Investment? An investment is a res...	1212
111	Quality and affordable healthcare is a vision...	847
112	Analytics is a statistical scientific process...	516
113	Big Data To begin with I shall first like to ...	733

	Polarity Score	Subjectivity Score	number_sentense
0	0.534389	0.000135	77
1	0.278969	0.000230	80
2	0.351348	0.000088	85
3	0.534389	0.000163	95
4	0.492383	0.000157	79
..	...	...	...
109	0.433331	0.000268	49
110	0.692302	0.000107	62
111	-0.005464	0.000216	66
112	0.954018	0.000337	29
113	0.388058	0.000366	66

	Average Sentense Length	Complex Word Count	Percentage of Complex words
0	18.13	798	57.16
1	12.65	397	39.23
2	14.76	707	56.33
3	12.18	513	44.34
4	15.86	627	

```
50.04
..                        ...                 ...
...
109                    13.71                 331
49.26
110                    19.55                 559
46.12
111                    12.83                 460
54.31
112                    17.79                 308
59.69
113                    11.11                 389
53.07

[114 rows x 14 columns]
```

Fog index

```
f_index = []
for i, row in tqdm(df.iterrows(), total=len(df)):
    sentense_length = row['Average Sentense Length']
    complex_word = row['Percentage of Complex words']
    fog_index = 0.4 * (sentense_length + complex_word)
    f_index.append(fog_index)
```

{"model_id":"6e1140f0cd0643a5b96581e4e0d380e6","version_major":2,"vers
ion_minor":0}

```
df['Fog Index'] = f_index
df
```

```
     Ids    neg    neu    pos  compound  \
0      0  0.044  0.811  0.145    0.9996
1      1  0.084  0.767  0.149    0.9984
2      2  0.036  0.889  0.075    0.9891
3      3  0.044  0.811  0.145    0.9995
4      4  0.050  0.804  0.147    0.9996
..   ...    ...    ...    ...       ...
109  109  0.051  0.820  0.129    0.9973
110  110  0.020  0.871  0.110    0.9993
111  111  0.092  0.818  0.091   -0.9301
112  112  0.004  0.825  0.170    0.9989
113  113  0.082  0.732  0.186    0.9990

                                              Websites  \
0    https://insights.blackcoffer.com/ai-in-healthc...
1    https://insights.blackcoffer.com/what-if-the-c...
2    https://insights.blackcoffer.com/what-jobs-wil...
3    https://insights.blackcoffer.com/will-machine-...
4    https://insights.blackcoffer.com/will-ai-repla...
..                                                 ...
```

```
109  https://insights.blackcoffer.com/blockchain-fo...
110  https://insights.blackcoffer.com/the-future-of...
111  https://insights.blackcoffer.com/big-data-anal...
112  https://insights.blackcoffer.com/business-anal...
113  https://insights.blackcoffer.com/challenges-an...

                                                Context  Count of words
\
0     Introduction “If anything kills over 10 milli...            1396

1     Human minds, a fascination in itself carrying...            1012

2     Introduction AI is rapidly evolving in the em...            1255

3     “Anything that could give rise to smarter-tha...            1157

4     “Machine intelligence is the last invention t...            1253

..                                                ...             ...

109   Reconciling with the financial realities of a...             672

110   What Is an Investment? An investment is a res...            1212

111   Quality and affordable healthcare is a vision...             847

112   Analytics is a statistical scientific process...             516

113   Big Data To begin with I shall first like to ...             733


     Polarity Score  Subjectivity Score  number_sentense  \
0          0.534389            0.000135               77
1          0.278969            0.000230               80
2          0.351348            0.000088               85
3          0.534389            0.000163               95
4          0.492383            0.000157               79
..              ...                 ...              ...
109        0.433331            0.000268               49
110        0.692302            0.000107               62
111       -0.005464            0.000216               66
112        0.954018            0.000337               29
113        0.388058            0.000366               66

     Average Sentense Length  Complex Word Count  Percentage of
Complex words  \
0                      18.13                 798
57.16
1                      12.65                 397
```

```
39.23
2                       14.76                  707
56.33
3                       12.18                  513
44.34
4                       15.86                  627
50.04
..                        ...                  ...
...
109                     13.71                  331
49.26
110                     19.55                  559
46.12
111                     12.83                  460
54.31
112                     17.79                  308
59.69
113                     11.11                  389
53.07

     Fog Index
0       30.116
1       20.752
2       28.436
3       22.608
4       26.360
..         ...
109     25.188
110     26.268
111     26.856
112     30.992
113     25.672

[114 rows x 15 columns]
```

##Words per sentense

```
word_per_sentense = []

for i, row in tqdm(df.iterrows(), total=len(df)):
  text = row['Context']
  total_sentense = sent_tokenize(text)
  sentense_count = len(total_sentense)

  words = []

  for sentense in total_sentense:
    for word in sentense:
      words.append(word)
```

```
  word_count = len(words)

  wps = round((word_count / sentense_count),2)
  word_per_sentense.append(wps)
```

{"model_id":"bbd0dc46db44412a83bfd7602552351e","version_major":2,"version_minor":0}

```
df['Average Number of Words Per Sentence'] = word_per_sentense
df
```

```
     Ids    neg    neu    pos  compound  \
0      0  0.044  0.811  0.145    0.9996
1      1  0.084  0.767  0.149    0.9984
2      2  0.036  0.889  0.075    0.9891
3      3  0.044  0.811  0.145    0.9995
4      4  0.050  0.804  0.147    0.9996
..   ...    ...    ...    ...       ...
109  109  0.051  0.820  0.129    0.9973
110  110  0.020  0.871  0.110    0.9993
111  111  0.092  0.818  0.091   -0.9301
112  112  0.004  0.825  0.170    0.9989
113  113  0.082  0.732  0.186    0.9990

                                              Websites  \
0    https://insights.blackcoffer.com/ai-in-healthc...
1    https://insights.blackcoffer.com/what-if-the-c...
2    https://insights.blackcoffer.com/what-jobs-wil...
3    https://insights.blackcoffer.com/will-machine-...
4    https://insights.blackcoffer.com/will-ai-repla...
..                                                 ...
109  https://insights.blackcoffer.com/blockchain-fo...
110  https://insights.blackcoffer.com/the-future-of...
111  https://insights.blackcoffer.com/big-data-anal...
112  https://insights.blackcoffer.com/business-anal...
113  https://insights.blackcoffer.com/challenges-an...

                                               Context  Count of words
\
0     Introduction “If anything kills over 10 milli...            1396

1     Human minds, a fascination in itself carrying...            1012

2     Introduction AI is rapidly evolving in the em...            1255

3     “Anything that could give rise to smarter-tha...            1157

4     “Machine intelligence is the last invention t...            1253

..                                                 ...             ...
```

```
109   Reconciling with the financial realities of a...             672

110   What Is an Investment? An investment is a res...            1212

111   Quality and affordable healthcare is a vision...             847

112   Analytics is a statistical scientific process...             516

113   Big Data To begin with I shall first like to ...             733


     Polarity Score  Subjectivity Score  number_sentense  \
0          0.534389            0.000135               77
1          0.278969            0.000230               80
2          0.351348            0.000088               85
3          0.534389            0.000163               95
4          0.492383            0.000157               79
..              ...                 ...              ...
109        0.433331            0.000268               49
110        0.692302            0.000107               62
111       -0.005464            0.000216               66
112        0.954018            0.000337               29
113        0.388058            0.000366               66

     Average Sentense Length  Complex Word Count  Percentage of
Complex words  \
0                      18.13                 798
57.16
1                      12.65                 397
39.23
2                      14.76                 707
56.33
3                      12.18                 513
44.34
4                      15.86                 627
50.04
..                       ...                 ...
...
109                    13.71                 331
49.26
110                    19.55                 559
46.12
111                    12.83                 460
54.31
112                    17.79                 308
59.69
113                    11.11                 389
53.07
```

```
     Fog Index  Average Number of Words Per Sentence
0       30.116                                157.13
1       20.752                                104.96
2       28.436                                129.80
3       22.608                                101.79
4       26.360                                136.63
..         ...                                   ...
109     25.188                                120.00
110     26.268                                156.87
111     26.856                                108.02
112     30.992                                167.21
113     25.672                                 96.35

[114 rows x 16 columns]
```

##Syllable per word

```
syllable_per_word = []

for i, row in tqdm(df.iterrows(), total=len(df)):
  text = row['Context']
  total_sentense = sent_tokenize(text)

  words = []
  sc = []

  for sentense in total_sentense:
    for word in sentense:
      words.append(word)

  for word in words:
    count = textstatistics().syllable_count(word)
    sc.append(count)

  spw = round((len(sc) / len(words)),2)
  syllable_per_word.append(spw)
```

{"model_id":"d5728104539c48749b8ea7777805c15b","version_major":2,"version_minor":0}

```
df['Syllable Count Per Word'] = syllable_per_word
df
```

```
     Ids    neg    neu    pos  compound  \
0      0  0.044  0.811  0.145    0.9996
1      1  0.084  0.767  0.149    0.9984
2      2  0.036  0.889  0.075    0.9891
3      3  0.044  0.811  0.145    0.9995
4      4  0.050  0.804  0.147    0.9996
..   ...    ...    ...    ...       ...
```

```
109  109  0.051  0.820  0.129    0.9973
110  110  0.020  0.871  0.110    0.9993
111  111  0.092  0.818  0.091   -0.9301
112  112  0.004  0.825  0.170    0.9989
113  113  0.082  0.732  0.186    0.9990

                                                Websites  \
0    https://insights.blackcoffer.com/ai-in-healthc...
1    https://insights.blackcoffer.com/what-if-the-c...
2    https://insights.blackcoffer.com/what-jobs-wil...
3    https://insights.blackcoffer.com/will-machine-...
4    https://insights.blackcoffer.com/will-ai-repla...
..                                                 ...
109  https://insights.blackcoffer.com/blockchain-fo...
110  https://insights.blackcoffer.com/the-future-of...
111  https://insights.blackcoffer.com/big-data-anal...
112  https://insights.blackcoffer.com/business-anal...
113  https://insights.blackcoffer.com/challenges-an...

                                                Context  Count of words
\
0     Introduction “If anything kills over 10 milli...            1396

1     Human minds, a fascination in itself carrying...            1012

2     Introduction AI is rapidly evolving in the em...            1255

3     “Anything that could give rise to smarter-tha...            1157

4     “Machine intelligence is the last invention t...            1253

..                                                 ...             ...

109   Reconciling with the financial realities of a...             672

110   What Is an Investment? An investment is a res...            1212

111   Quality and affordable healthcare is a vision...             847

112   Analytics is a statistical scientific process...             516

113   Big Data To begin with I shall first like to ...             733


     Polarity Score  Subjectivity Score  number_sentense  \
0          0.534389            0.000135               77
1          0.278969            0.000230               80
2          0.351348            0.000088               85
3          0.534389            0.000163               95
```

```
4          0.492383            0.000157               79
..              ...                 ...              ...
109        0.433331            0.000268               49
110        0.692302            0.000107               62
111       -0.005464            0.000216               66
112        0.954018            0.000337               29
113        0.388058            0.000366               66

     Average Sentense Length  Complex Word Count  Percentage of
Complex words  \
0                       18.13                 798
57.16
1                       12.65                 397
39.23
2                       14.76                 707
56.33
3                       12.18                 513
44.34
4                       15.86                 627
50.04
..                        ...                 ...
...
109                     13.71                 331
49.26
110                     19.55                 559
46.12
111                     12.83                 460
54.31
112                     17.79                 308
59.69
113                     11.11                 389
53.07

     Fog Index  Average Number of Words Per Sentence  Syllable Count
Per Word
0       30.116                                157.13
1.0
1       20.752                                104.96
1.0
2       28.436                                129.80
1.0
3       22.608                                101.79
1.0
4       26.360                                136.63
1.0
..         ...                                   ...
...
109     25.188                                120.00
1.0
110     26.268                                156.87
```

```
1.0
111     26.856                                    108.02
1.0
112     30.992                                    167.21
1.0
113     25.672                                     96.35
1.0

[114 rows x 17 columns]
```

##Personal pronouns

```
import re

personal_pronouns_count = []

for i, row in tqdm(df.iterrows(), total=len(df)):
  text = row['Context']

  pronounRegex = re.compile(r'I|we|my|ours|us',re.I)
  pronouns = pronounRegex.findall(text)
  personal_pronouns_count.append(len(pronouns))
```

{"model_id":"2259da3ef361491d88def402cbf0d986","version_major":2,"version_minor":0}

```
df['Personal Pronouns'] = personal_pronouns_count
df
```

```
     Ids    neg    neu    pos  compound  \
0      0  0.044  0.811  0.145    0.9996
1      1  0.084  0.767  0.149    0.9984
2      2  0.036  0.889  0.075    0.9891
3      3  0.044  0.811  0.145    0.9995
4      4  0.050  0.804  0.147    0.9996
..   ...    ...    ...    ...       ...
109  109  0.051  0.820  0.129    0.9973
110  110  0.020  0.871  0.110    0.9993
111  111  0.092  0.818  0.091   -0.9301
112  112  0.004  0.825  0.170    0.9989
113  113  0.082  0.732  0.186    0.9990

                                              Websites  \
0    https://insights.blackcoffer.com/ai-in-healthc...
1    https://insights.blackcoffer.com/what-if-the-c...
2    https://insights.blackcoffer.com/what-jobs-wil...
3    https://insights.blackcoffer.com/will-machine-...
4    https://insights.blackcoffer.com/will-ai-repla...
..                                                 ...
109  https://insights.blackcoffer.com/blockchain-fo...
110  https://insights.blackcoffer.com/the-future-of...
```

```
111  https://insights.blackcoffer.com/big-data-anal...
112  https://insights.blackcoffer.com/business-anal...
113  https://insights.blackcoffer.com/challenges-an...

                                                Context  Count of words
\
0     Introduction “If anything kills over 10 milli...            1396

1     Human minds, a fascination in itself carrying...            1012

2     Introduction AI is rapidly evolving in the em...            1255

3     “Anything that could give rise to smarter-tha...            1157

4     “Machine intelligence is the last invention t...            1253

..                                                ...             ...

109   Reconciling with the financial realities of a...             672

110   What Is an Investment? An investment is a res...            1212

111   Quality and affordable healthcare is a vision...             847

112   Analytics is a statistical scientific process...             516

113   Big Data To begin with I shall first like to ...             733


     Polarity Score  Subjectivity Score  number_sentense  \
0          0.534389            0.000135               77
1          0.278969            0.000230               80
2          0.351348            0.000088               85
3          0.534389            0.000163               95
4          0.492383            0.000157               79
..              ...                 ...              ...
109        0.433331            0.000268               49
110        0.692302            0.000107               62
111       -0.005464            0.000216               66
112        0.954018            0.000337               29
113        0.388058            0.000366               66

     Average Sentense Length  Complex Word Count  Percentage of
Complex words  \
0                      18.13                 798
57.16
1                      12.65                 397
39.23
2                      14.76                 707
```

```
56.33
3                      12.18                 513
44.34
4                      15.86                 627
50.04
..                       ...                 ...
...
109                    13.71                 331
49.26
110                    19.55                 559
46.12
111                    12.83                 460
54.31
112                    17.79                 308
59.69
113                    11.11                 389
53.07

     Fog Index  Average Number of Words Per Sentence  Syllable Count
Per Word  \
0       30.116                                157.13
1.0
1       20.752                                104.96
1.0
2       28.436                                129.80
1.0
3       22.608                                101.79
1.0
4       26.360                                136.63
1.0
..         ...                                   ...
...
109     25.188                                120.00
1.0
110     26.268                                156.87
1.0
111     26.856                                108.02
1.0
112     30.992                                167.21
1.0
113     25.672                                 96.35
1.0

     Personal Pronouns
0                  933
1                  600
2                  875
3                  684
4                  760
..                 ...
```

```
109                402
110                677
111                518
112                397
113                467

[114 rows x 18 columns]
```

##Average word length

```
average_length = []

for i, row in tqdm(df.iterrows(), total=len(df)):
  text = row['Context']
  total_sentense = sent_tokenize(text)

  words = []
  sc = []

  for sentense in total_sentense:
    for word in sentense:
      words.append(word)

  average = round(sum(len(word) for word in words) / len(words),2)
  average_length.append(average)
```

{"model_id":"53c57dd217934ebb8f99f85390f0e251","version_major":2,"version_minor":0}

```
df['Average Word Length'] = average_length
df
```

```
     Ids    neg    neu    pos  compound  \
0      0  0.044  0.811  0.145    0.9996
1      1  0.084  0.767  0.149    0.9984
2      2  0.036  0.889  0.075    0.9891
3      3  0.044  0.811  0.145    0.9995
4      4  0.050  0.804  0.147    0.9996
..   ...    ...    ...    ...       ...
109  109  0.051  0.820  0.129    0.9973
110  110  0.020  0.871  0.110    0.9993
111  111  0.092  0.818  0.091   -0.9301
112  112  0.004  0.825  0.170    0.9989
113  113  0.082  0.732  0.186    0.9990

                                              Websites  \
0    https://insights.blackcoffer.com/ai-in-healthc...
1    https://insights.blackcoffer.com/what-if-the-c...
2    https://insights.blackcoffer.com/what-jobs-wil...
3    https://insights.blackcoffer.com/will-machine-...
4    https://insights.blackcoffer.com/will-ai-repla...
```

```
..                                                  ...
109  https://insights.blackcoffer.com/blockchain-fo...
110  https://insights.blackcoffer.com/the-future-of...
111  https://insights.blackcoffer.com/big-data-anal...
112  https://insights.blackcoffer.com/business-anal...
113  https://insights.blackcoffer.com/challenges-an...

                                                Context  Count of words
\
0     Introduction “If anything kills over 10 milli...            1396

1     Human minds, a fascination in itself carrying...            1012

2     Introduction AI is rapidly evolving in the em...            1255

3     “Anything that could give rise to smarter-tha...            1157

4     “Machine intelligence is the last invention t...            1253

..                                                ...             ...

109   Reconciling with the financial realities of a...             672

110   What Is an Investment? An investment is a res...            1212

111   Quality and affordable healthcare is a vision...             847

112   Analytics is a statistical scientific process...             516

113   Big Data To begin with I shall first like to ...             733


     Polarity Score  Subjectivity Score  number_sentense  \
0          0.534389            0.000135               77
1          0.278969            0.000230               80
2          0.351348            0.000088               85
3          0.534389            0.000163               95
4          0.492383            0.000157               79
..              ...                 ...              ...
109        0.433331            0.000268               49
110        0.692302            0.000107               62
111       -0.005464            0.000216               66
112        0.954018            0.000337               29
113        0.388058            0.000366               66

     Average Sentense Length  Complex Word Count  Percentage of
Complex words  \
0                      18.13                 798
57.16
```

```
1                     12.65                 397
39.23
2                     14.76                 707
56.33
3                     12.18                 513
44.34
4                     15.86                 627
50.04
..                      ...                 ...
...
109                   13.71                 331
49.26
110                   19.55                 559
46.12
111                   12.83                 460
54.31
112                   17.79                 308
59.69
113                   11.11                 389
53.07

     Fog Index  Average Number of Words Per Sentence  Syllable Count
Per Word  \
0       30.116                                157.13
1.0
1       20.752                                104.96
1.0
2       28.436                                129.80
1.0
3       22.608                                101.79
1.0
4       26.360                                136.63
1.0
..         ...                                   ...
...
109     25.188                                120.00
1.0
110     26.268                                156.87
1.0
111     26.856                                108.02
1.0
112     30.992                                167.21
1.0
113     25.672                                 96.35
1.0

     Personal Pronouns  Average Word Length
0                  933                  1.0
1                  600                  1.0
2                  875                  1.0
```

```
3                  684                  1.0
4                  760                  1.0
..                 ...                  ...
109                402                  1.0
110                677                  1.0
111                518                  1.0
112                397                  1.0
113                467                  1.0

[114 rows x 19 columns]
```

###Renaming columns

```
df = df.rename(columns={'neg':"Negative Score",'pos':"Positive
Score",'Ids':"URL ID",'Websites':"URL",'Count of words':"Word
Count",'Syllable Count Per Word':"Syllable Per Word"})

df.head(1)
```

```
   URL ID  Negative Score    neu  Positive Score  compound  \
0       0           0.044  0.811           0.145    0.9996

                                                 URL  \
0  https://insights.blackcoffer.com/ai-in-healthc...

                                             Context  Word Count  \
0   Introduction “If anything kills over 10 milli...        1396

   Polarity Score  Subjectivity Score  number_sentense  \
0        0.534389            0.000135               77

   Average Sentense Length  Complex Word Count  Percentage of Complex
words  \
0                    18.13                 798
57.16

   Fog Index  Average Number of Words Per Sentence  Syllable Per Word
\
0     30.116                                157.13                1.0


   Personal Pronouns  Average Word Length
0                933                  1.0
```

###Dropping unnecessary columns

```
df = df.drop(['neu','compound','Context','number_sentense'],axis = 1)
df.head(1)
```

```
   URL ID  Negative Score  Positive Score  \
0       0           0.044           0.145
```

```
                                                 URL  Word Count  \
0  https://insights.blackcoffer.com/ai-in-healthc...        1396

   Polarity Score  Subjectivity Score  Average Sentense Length  \
0        0.534389            0.000135                    18.13

   Complex Word Count  Percentage of Complex words  Fog Index  \
0                 798                        57.16     30.116

   Average Number of Words Per Sentence  Syllable Per Word  Personal
Pronouns  \
0                                157.13                1.0
933

   Average Word Length
0                  1.0
```

```
df.columns
```

```
Index(['URL ID', 'Negative Score', 'neu', 'Positive Score',
'compound', 'URL',
       'Context', 'Word Count', 'Polarity Score', 'Subjectivity
Score',
       'number_sentense', 'Average Sentense Length', 'Complex Word
Count',
       'Percentage of Complex words', 'Fog Index',
       'Average Number of Words Per Sentence', 'Syllable Per Word',
       'Personal Pronouns', 'Average Word Length'],
      dtype='object')
```

###Rearranging the columns

```
df  = df[['URL ID','URL','Positive Score','Negative Score','Polarity
Score','Subjectivity Score','Average Sentense Length','Percentage of
Complex words',
          'Fog Index','Average Number of Words Per Sentence','Complex
Word Count','Word Count','Syllable Per Word','Personal Pronouns',
'Average Word Length']]
```

```
df.head(2)
```

```
   URL ID                                                URL  Positive
Score  \
0       0  https://insights.blackcoffer.com/ai-in-healthc...
0.145
1       1  https://insights.blackcoffer.com/what-if-the-c...
0.149

   Negative Score  Polarity Score  Subjectivity Score  \
0           0.044        0.534389            0.000135
```

```
1           0.084        0.278969            0.000230

   Average Sentense Length  Percentage of Complex words  Fog Index  \
0                    18.13                        57.16     30.116
1                    12.65                        39.23     20.752

   Average Number of Words Per Sentence  Complex Word Count  Word
Count  \
0                                157.13                 798
1396
1                                104.96                 397
1012

   Syllable Per Word  Personal Pronouns  Average Word Length
0                1.0                933                  1.0
1                1.0                600                  1.0
```

###Saving the Final Output to an excel file

```
df.to_excel('Final_Output.xlsx')
```