```
In [4]:   import pandas as pd
          import numpy as np
          import re
```

Reading the input excel file

```
In [5]:   df = pd.read_excel("Input.xlsx")
          df.head()
```

Out[5]:

	URL_ID	URL
0	37	https://insights.blackcoffer.com/ai-in-healthc...
1	38	https://insights.blackcoffer.com/what-if-the-c...
2	39	https://insights.blackcoffer.com/what-jobs-wil...
3	40	https://insights.blackcoffer.com/will-machine-...
4	41	https://insights.blackcoffer.com/will-ai-repla...

```
In [6]:   links =[]
          for i in df['URL']:
              links.append(i)
```

```
In [3]:   #links
          #len(links)
```

Download the required libraries

```
In [5]:   #!pip install bs4 requests html5lib
          #!pip install textblob
```

```
In [7]:   from urllib.request import urlopen
          from bs4 import BeautifulSoup
          import requests
          from textblob import TextBlob
```

Getting all the titles of the website links ....

```
In [9]:   all_headers = []
          def links_process1(url):

              # using the BeautifulSoup module
              r = requests.get(url)
              if r.status_code == 200 :
                  soup = BeautifulSoup (r.content , 'html.parser')
                  all_headers.append(soup.title.get_text())

              else :
                  all_headers.append("None")
```

```
In [10]:   for i in links:
               links_process1(i)
```

```
In [74]:   #all_headers
```

Creating a new dataframe

```
In [11]:   df2 = df
           df2.head()
```

Out[11]:

	URL_ID	URL
0	37	https://insights.blackcoffer.com/ai-in-healthc...
1	38	https://insights.blackcoffer.com/what-if-the-c...
2	39	https://insights.blackcoffer.com/what-jobs-wil...
3	40	https://insights.blackcoffer.com/will-machine-...
4	41	https://insights.blackcoffer.com/will-ai-repla...

```
In [12]:   df2["titles"] = all_headers
           df2.head()
```

Out[12]:

	URL_ID	URL	titles
0	37	https://insights.blackcoffer.com/ai-in-healthc...	AI in healthcare to Improve Patient Outcomes |...
1	38	https://insights.blackcoffer.com/what-if-the-c...	What if the Creation is Taking Over the Creato...
2	39	https://insights.blackcoffer.com/what-jobs-wil...	What Jobs Will Robots Take From Humans in The ...
3	40	https://insights.blackcoffer.com/will-machine-...	Will Machine Replace The Human in the Future o...
4	41	https://insights.blackcoffer.com/will-ai-repla...	Will AI Replace Us or Work With Us? | Blackcof...

Find all the paragraphs

In [13]:
```python
# creating a handler function to save info

cont = []

def save_data (url ):

    r = requests.get(url)
    if r.status_code == 200:

        soup = BeautifulSoup(r.content , "html.parser")
        p_tree = soup.find("div" , class_ = "td-post-content")

        #l_text = []
        text = ""
        #print(type(p_tree))
        pts = p_tree.find_all('p')

        for x in pts:

            text += x.get_text()

        cont.append(text)


    else:
        cont.append("No Content ")
```

In [14]:
```python
for l in links:
    save_data(l)
```

In [15]:
```python
#Len(cont)
```

In [16]:
```python
df2["web-content"] = cont
df2.head()
```

Out[16]:

	URL_ID	URL	titles	web-content
0	37	https://insights.blackcoffer.com/ai-in-healthc...	AI in healthcare to Improve Patient Outcomes |...	Introduction"If anything kills over 10 million...
1	38	https://insights.blackcoffer.com/what-if-the-c...	What if the Creation is Taking Over the Creato...	Human minds, a fascination in itself carrying ...
2	39	https://insights.blackcoffer.com/what-jobs-wil...	What Jobs Will Robots Take From Humans in The ...	IntroductionAI is rapidly evolving in the empl...
3	40	https://insights.blackcoffer.com/will-machine-...	Will Machine Replace The Human in the Future o...	"Anything that could give rise to smarter-than...
4	41	https://insights.blackcoffer.com/will-ai-repla...	Will AI Replace Us or Work With Us? | Blackcof...	"Machine intelligence is the last invention th...

Creating a directory "text_data" to store the extract text data into it

```
In [87]: import os
         path = "./"
         #c_path = os.listdir(path)
         file_path = os.path.join( path , "text_data")
```

Saving the files as per the name of the url_id

```
In [26]: names = df["URL_ID"]
         names = names.copy()
         names
```

```
Out[26]: 0       37
         1       38
         2       39
         3       40
         4       41
                ...
         109    146
         110    147
         111    148
         112    149
         113    150
         Name: URL_ID, Length: 114, dtype: int64
```

Saving the web-content text data into the seprate files

```
In [27]: text_file_path = []

for y in range(0 , 114):

    c_text = str(cont[y])


    y = os.path.join(  file_path , str(names[y]))
    print(y)
    text_file_path.append(y)


    with open(y, "w" ,encoding='utf8') as file:

        file.write(c_text)
        file.close()
```

```
./text_data\131
./text_data\132
./text_data\133
./text_data\134
./text_data\135
./text_data\136
./text_data\137
./text_data\138
./text_data\139
./text_data\140
./text_data\141
./text_data\142
./text_data\143
./text_data\144
./text_data\145
./text_data\146
./text_data\147
./text_data\148
./text_data\149
./text_data\150
```

```
In [32]: df2["text_file_paths"] = text_file_path
df2.head()

df2.to_csv("data1.csv")
```

Step 2 - Sentiment Analysis

Performing analysis on web-content

In [33]:
```python
df3 = pd.read_csv("data1.csv")
df3.head()
```

Out[33]:

	Unnamed: 0	URL_ID	URL	titles	web-content	text_file_paths
0	0	37	https://insights.blackcoffer.com/ai-in-healthc...	AI in healthcare to Improve Patient Outcomes |...	Introduction"If anything kills over 10 million...	./text_data\37
1	1	38	https://insights.blackcoffer.com/what-if-the-c...	What if the Creation is Taking Over the Creato...	Human minds, a fascination in itself carrying ...	./text_data\38
2	2	39	https://insights.blackcoffer.com/what-jobs-wil...	What Jobs Will Robots Take From Humans in The ...	IntroductionAI is rapidly evolving in the empl...	./text_data\39
3	3	40	https://insights.blackcoffer.com/will-machine-...	Will Machine Replace The Human in the Future o...	"Anything that could give rise to smarter-than...	./text_data\40

In [34]:
```python
df3.columns
```

Out[34]:
```
Index(['Unnamed: 0', 'URL_ID', 'URL', 'titles', 'web-content',
       'text_file_paths'],
      dtype='object')
```

In [35]:
```python
df3 = df3.drop('Unnamed: 0', axis = 1)
```

In [36]:
```python
df3.head()
```

Out[36]:

	URL_ID	URL	titles	web-content	text_file_paths
0	37	https://insights.blackcoffer.com/ai-in-healthc...	AI in healthcare to Improve Patient Outcomes |...	Introduction"If anything kills over 10 million...	./text_data\37
1	38	https://insights.blackcoffer.com/what-if-the-c...	What if the Creation is Taking Over the Creato...	Human minds, a fascination in itself carrying ...	./text_data\38
2	39	https://insights.blackcoffer.com/what-jobs-wil...	What Jobs Will Robots Take From Humans in The ...	IntroductionAI is rapidly evolving in the empl...	./text_data\39
3	40	https://insights.blackcoffer.com/will-machine-...	Will Machine Replace The Human in the Future o...	"Anything that could give rise to smarter-than...	./text_data\40
4	41	https://insights.blackcoffer.com/will-ai-repla...	Will AI Replace Us or Work With Us? | Blackcof...	"Machine intelligence is the last invention th...	./text_data\41

Word Counts

In [37]:
```python
counter = lambda x: len(x.split())
df3["WORD_COUNT"] =  df3["web-content"].apply(counter)
df3.head()
```

Out[37]:

	URL_ID	URL	titles	web-content	text_file_paths	WORD_COUNT
0	37	https://insights.blackcoffer.com/ai-in-healthc...	AI in healthcare to Improve Patient Outcomes |...	Introduction"If anything kills over 10 million...	./text_data\37	1745
1	38	https://insights.blackcoffer.com/what-if-the-c...	What if the Creation is Taking Over the Creato...	Human minds, a fascination in itself carrying ...	./text_data\38	1389
2	39	https://insights.blackcoffer.com/what-jobs-wil...	What Jobs Will Robots Take From Humans in The ...	IntroductionAI is rapidly evolving in the empl...	./text_data\39	1661
3	40	https://insights.blackcoffer.com/will-machine-...	Will Machine Replace The Human in the Future o...	"Anything that could give rise to smarter-than...	./text_data\40	1562
4	41	https://insights.blackcoffer.com/will-ai-repla...	Will AI Replace Us or Work With Us? | Blackcof...	"Machine intelligence is the last invention th...	./text_data\41	1676

Senetence count

```
In [56]: sentence_counter = lambda x: len(x.split("."))
         df3["SENTENCE_COUNT"] = df3["web-content"].apply(sentence_counter)
         df3.head()
```

Out[56]:

	URL_ID	URL	titles	web-content	text_file_paths	WORD_COUNT	SENTENC
0	37	https://insights.blackcoffer.com/ai-in-healthc...	AI in healthcare to Improve Patient Outcomes |...	Introduction"If anything kills over 10 million...	./text_data\37	1745	
1	38	https://insights.blackcoffer.com/what-if-the-c...	What if the Creation is Taking Over the Creato...	Human minds, a fascination in itself carrying ...	./text_data\38	1389	
2	39	https://insights.blackcoffer.com/what-jobs-wil...	What Jobs Will Robots Take From Humans in The ...	IntroductionAI is rapidly evolving in the empl...	./text_data\39	1661	
3	40	https://insights.blackcoffer.com/will-machine-...	Will Machine Replace The Human in the Future o...	"Anything that could give rise to smarter-than...	./text_data\40	1562	
4	41	https://insights.blackcoffer.com/will-ai-repla...	Will AI Replace Us or Work With Us? | Blackcof...	"Machine intelligence is the last invention th...	./text_data\41	1676	

Functions

```
In [57]: def avg_words(x):
             words = x.split()
             return sum(len(word) for word in words) / len (words)
```

```
In [58]: def avg_sen_len(x):
             words = x.split(".")
             return sum(len(word) for word in words) / len (words)
```

In [59]:
```python
def count_syllables(word):

    return float(len(
        re.findall('(?!e$)[aeiouy]+', word, re.I) +
        re.findall('^[^aeiouy]*e$', word, re.I)
    )/len(word))

#count_syllables("test")
```

In [62]:
```python
pronounRegex = re.compile(r'\b(I|we|my|ours|(?-i:us))\b',re.I)
pronouns = pronounRegex.findall(' I me  you We and all of us make this team tweek, he is f
#print(len(pronouns))
```

In [63]:
```python
def pronoun_counter(x):
    pronounRegex = re.compile(r'\b(I|we|my|ours|(?-i:us))\b',re.I)
    pronouns = pronounRegex.findall(x)
    return len(pronouns)
```

In [64]:
```python
def syllable_counter(x):
    words = x.split()
    return sum(count_syllables(word) for word in words) / len (words)
```

In [65]:
```python
df3['SYLLABLE_PER_WORD'] = df3["web-content"].apply(lambda x: syllable_counter(x) )
df3["AVERAGE_SENTENCE_LENGTH"] = df3["web-content"].apply(lambda x: avg_sen_len(x))
df3["AVERAGE_WORDS"] = df3["web-content"].apply(lambda x: avg_words(x))
df3['PERSONAL_PRONOUNS'] = df3["web-content"].apply(lambda x: pronoun_counter(x) )
```

In [46]:
```python
## Finding out the complex words
### Removing simple words result out having complex words
```

In [66]:
```python
#!pip install nltk
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
```

```
[nltk_data] Downloading package stopwords to C:\Users\ACER ES1
[nltk_data]     512\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

In [67]:
```python
stop_words = stopwords.words('english')
#len(stop_words)
```

In [75]:
```python
df3['COMPLEX_WORDS'] = df3["web-content"].apply(lambda x: len([word for word in x.split()
df3['PER_COMPLEX_WORDS'] = df3["web-content"].apply(lambda x: (len([word for word in x.spl
```

In [69]:
```python
from textblob import TextBlob
```

In [70]:
```python
df3["POLARITY_SCORE"] = df3["web-content"].apply(lambda x: TextBlob(x).sentiment[0])
df3["SUBJECTIVITY_SCORE"] = df3["web-content"].apply(lambda x: TextBlob(x).sentiment[1])
#df.head()
```

In [71]:
```python
def positive_score(text):
    score = 0

    score = TextBlob(text).sentiment[0]

    if score > 0 :
        return score
    else:
        return 0
```

In [72]:
```python
def negative_score(text):
    score = 0

    score = TextBlob(text).sentiment[0]

    if score < 0 :
        return score
    else:
        return 0
```

In [73]:
```python
df3["NEGATIVE_SCORE"] = df3["web-content"].apply(lambda x: negative_score(x))
df3["POSITIVE_SCORE"] = df3["web-content"].apply(lambda x: positive_score(x))
```

In [76]:
```python
# Calculate the fog index
#fog = 0.4 * ((word_count / len(sentence)) )

df3["FOG_INDEX"] = 0.4* (( df3['WORD_COUNT'] / df3['SENTENCE_COUNT']) + df3['COMPLEX_WORDS
```

In [77]: `df3.head()`

Out[77]:

	URL_ID	URL	titles	web-content	text_file_paths	WORD_COUNT	SENTENC
0	37	https://insights.blackcoffer.com/ai-in-healthc...	AI in healthcare to Improve Patient Outcomes |...	Introduction"If anything kills over 10 million...	./text_data\37	1745	
1	38	https://insights.blackcoffer.com/what-if-the-c...	What if the Creation is Taking Over the Creato...	Human minds, a fascination in itself carrying ...	./text_data\38	1389	
2	39	https://insights.blackcoffer.com/what-jobs-wil...	What Jobs Will Robots Take From Humans in The ...	IntroductionAI is rapidly evolving in the empl...	./text_data\39	1661	
3	40	https://insights.blackcoffer.com/will-machine-...	Will Machine Replace The Human in the Future o...	"Anything that could give rise to smarter-than...	./text_data\40	1562	
4	41	https://insights.blackcoffer.com/will-ai-repla...	Will AI Replace Us or Work With Us? | Blackcof...	"Machine intelligence is the last invention th...	./text_data\41	1676	

In [78]: `df3.columns`

Out[78]:
```
Index(['URL_ID', 'URL', 'titles', 'web-content', 'text_file_paths',
       'WORD_COUNT', 'SENTENCE_COUNT', 'SYLLABLE_PER_WORD',
       'AVERAGE_SENTENCE_LENGTH', 'AVERAGE_WORDS', 'PERSONAL_PRONOUNS',
       'PER_COMPLEX_WORDS', 'POLARITY_SCORE', 'SUBJECTIVITY_SCORE',
       'NEGATIVE_SCORE', 'POSITIVE_SCORE', 'COMPLEX_WORDS', 'FOG_INDEX'],
      dtype='object')
```

In [80]: `df3 = df3.drop(['titles', 'web-content', 'text_file_paths'] , axis = 1)`

In [81]: `len(df3.columns)`

Out[81]: 15

Loading the o/p template

```
In [82]: op = pd.read_excel("output.xlsx")
         op.head()
```

Out[82]:

	URL_ID	URL	POSITIVE SCORE	NEGATIVE SCORE	POLARITY SCORE	SUBJECTIVITY SCORE	AVG SENTENCE LENGTH	PEI OF
0	37	https://insights.blackcoffer.com/ai-in-healthc...	NaN	NaN	NaN	NaN	NaN	
1	38	https://insights.blackcoffer.com/what-if-the-c...	NaN	NaN	NaN	NaN	NaN	
2	39	https://insights.blackcoffer.com/what-jobs-wil...	NaN	NaN	NaN	NaN	NaN	
3	40	https://insights.blackcoffer.com/will-machine-...	NaN	NaN	NaN	NaN	NaN	
4	41	https://insights.blackcoffer.com/will-ai-repla...	NaN	NaN	NaN	NaN	NaN	

```
In [83]: op_cols= op.columns
         cols = df3.columns
```

```
In [84]: op["POSITIVE SCORE"] = df3["POSITIVE_SCORE"]
         op["NEGATIVE SCORE"] = df3["NEGATIVE_SCORE"]
         op["POLARITY SCORE"] = df3["POLARITY_SCORE"]
         op["SUBJECTIVITY SCORE"] = df3["SUBJECTIVITY_SCORE"]
         op["AVG SENTENCE LENGTH"] = df3["AVERAGE_SENTENCE_LENGTH"]
         op["PERCENTAGE OF COMPLEX WORDS"] = df3["PER_COMPLEX_WORDS"]
         op["FOG INDEX"] = df3["FOG_INDEX"]
         op["AVG NUMBER OF WORDS PER SENTENCE"] = df3["AVERAGE_WORDS"]
         op["COMPLEX WORD COUNT"] = df3["COMPLEX_WORDS"]
         op["WORD COUNT"] = df3["POSITIVE_SCORE"]
         op["SYLLABLE PER WORD"] = df3["WORD_COUNT"]
         op["PERSONAL PRONOUNS"] = df3["PERSONAL_PRONOUNS"]
         op["AVG WORD LENGTH"] = df3["AVERAGE_WORDS"]
```

In [85]: `op.head()`

Out[85]:

	URL_ID	URL	POSITIVE SCORE	NEGATIVE SCORE	POLARITY SCORE	SUBJECTIVITY SCORE	AVG SENTENCE LENGTH	PER OF
0	37	https://insights.blackcoffer.com/ai-in-healthc...	0.138662	0.0	0.138662	0.467709	155.181818	
1	38	https://insights.blackcoffer.com/what-if-the-c...	0.084251	0.0	0.084251	0.422497	118.857143	
2	39	https://insights.blackcoffer.com/what-jobs-wil...	0.115244	0.0	0.115244	0.490544	128.070588	
3	40	https://insights.blackcoffer.com/will-machine-...	0.157602	0.0	0.157602	0.493522	111.819277	
4	41	https://insights.blackcoffer.com/will-ai-repla...	0.054676	0.0	0.054676	0.510968	128.629630	

Saving the final_results in final_op

In [86]: `op.to_excel("final_op.xlsx")`

In [ ]: