

```
In [1]: import nltk
import re
import pandas as pd

import os
from nltk.tokenize import word_tokenize, sent_tokenize
```

```
In [2]: nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\arnav\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

```
Out[2]: True
```

```
In [3]: folder_path='StopWords'
text_files = [f for f in os.listdir(folder_path) if f.endswith(".txt")]
```

```
In [4]: stopword_dic=[]
for file_name in text_files:
    file_path = os.path.join(folder_path, file_name)
    with open(file_path, 'r') as file:
        stopword_dic = set(file.read().split())
```

```
In [5]: Pos_dict_path=os.path.join('MasterDictionary', 'positive-words.txt')
Neg_dict_path=os.path.join('MasterDictionary', 'negative-words.txt')
```

```
In [6]: negative_word_dic=[]
with open(Neg_dict_path, 'r') as file:
    negative_word_dic = set(file.read().split())
```

```
In [7]: negative_word_dic
```

```
Out[7]: {'extremists',
'dangerous',
'freak',
'maniac',
'malicious',
'unwell',
'spitefully',
'anger',
'fooled',
'ineloquent',
'dissenter',
'freezing',
'insurrection',
'leery',
'interrupt',
'noxious',
'crooked',
'declining',
'remorselessness',
'bid...'
```

```
In [8]: positive_word_dic=[]  
        with open(Pos_dict_path, 'r') as file:  
            positive_word_dic = set(file.read().split())
```

```
In [9]: positive_word_dic
```

```
Out[9]: {'fearless',  
         'peaceable',  
         'favour',  
         'wonderful',  
         'golden',  
         'detachable',  
         'merriment',  
         'personages',  
         'upscale',  
         'adroitly',  
         'comfortable',  
         'inspiration',  
         'luxurious',  
         'modesty',  
         'celebratory',  
         'rightful',  
         'dexterous',  
         'energy-efficient',  
         'sustainable',  
         ...}
```

```
In [10]: def clean_text(text):  
         words = word_tokenize(text)  
         cleaned_words = [word.lower() for word in words if word.isalpha() and w  
         return cleaned_words
```

```
In [11]: def calculate_positive_score(text, positive_dict):  
         positive_words = [word for word in text if word in positive_dict]  
         return len(positive_words)
```

```
In [12]: def calculate_negative_score(text, negative_dict):  
         negative_words = [word for word in text if word in negative_dict]  
         return len(negative_words) * -1
```

```
In [13]: def calculate_polarity_score(positive_score, negative_score):  
         return (positive_score - negative_score) / (positive_score + negative_s
```

```
In [14]: def calculate_subjectivity_score(positive_score, negative_score, total_word  
         return (positive_score + negative_score) / (total_words + 0.000001)
```

```
In [15]: def analyze_readability(text):
    sentences = sent_tokenize(text)
    total_words = len(clean_text(text))
    average_sentence_length = total_words / len(sentences)

    complex_words = [word for word in clean_text(text) if syllable_count(word) > 3]
    percentage_complex_words = len(complex_words) / total_words
    fog_index = 0.4 * (average_sentence_length + percentage_complex_words)

    average_words_per_sentence = total_words / len(sentences)
    complex_word_count = len(complex_words)

    return average_sentence_length, percentage_complex_words, fog_index, av
```

```
In [16]: def syllable_count(word):
    vowels = "aeiouy"
    count = 0

    # Handle words ending with "es" and "ed"
    if word.endswith(("es", "ed")):
        pass
    else:
        for char in word:
            if char.lower() in vowels:
                count += 1

    return count
```

```
In [17]: def calculate_syllables_per_word(text):
    words = clean_text(text)
    syllables = sum(syllable_count(word) for word in words)
    return syllables / len(words)
```

```
In [18]: def count_personal_pronouns(text):
    personal_pronouns = re.findall(r'\b(?:I|we|my|ours|us)\b', text, flags=
    return len(personal_pronouns)
```

```
In [19]: def calculate_average_word_length(text):
    words = clean_text(text)
    total_characters = sum(len(word) for word in words)
    return total_characters / len(words)
```

```
In [20]: output_data = []
```

```
In [21]: input_data=[]  
         input_path=os.path.join('Extracted Data','blackassign0001')  
         with open(input_path, 'r') as file:  
             input_data= file.read()  
         input_data
```

Out[21]: 'ML and AI-based insurance premium model to predict premium to be charged by the insurance company\n\nWe have seen a huge development and dependence of people on technology in recent years. We have also seen the development of AI and ChatGPT in recent years. So it is a normal thing that we will be come fully dependent on technology by 2040. Information technology will be a major power for all the developing nations. As a member of a developing nation, India is rapidly growing its IT base. It has also grown some IT cities which will be the major control centres for Information technology by 2040.\n\nRising IT cities\n\nNoida:- Noida in Uttar Pradesh near New Delhi is an emerging IT sector now. Many large companies like Google, Microsoft, IBM, Infosys and others have set up their companies here. Noida has a market base of billions of dollars and is doing a great job of boosting the national economy. The establishment of so many software companies has made Noida an information technology hub.\n\nGurgaon:- Gurgaon in Haryana is also an emerging IT hub. Many large companies like Google, Microsoft, IBM, Infosys and others have set up their companies here. Gurgaon has a market base of billions of dollars and is doing a great job of boosting the national economy.\n\nBengaluru:- Bengaluru is called as the IT hub of India. It is also a smart city. Many large companies like Google, Microsoft, IBM, Infosys and others have set up their companies here. Bengaluru has a market base of billions of dollars and is doing a great job of boosting the national economy.\n\nKolkata:- Kolkata in West Bengal is an emerging major IT hub. The new Kolkata i.e. Saltlake Sector 5, New town, Rajarhat area of Kolkata is a major IT hub. The government is giving the software companies land at almost free of cost to set up the companies there. Many large companies like Google, Microsoft, IBM, Infosys and others have set up their companies here. Kolkata has a market base of billions of dollars and is doing a great job of boosting the national economy.\n\nImpact on Economy\n\nThere is a huge impact of the rising IT cities on our economy. Some of the effects are-\n\nDemand:- The rising IT cities will greatly help to boost our economy. These will create a huge demand for raw materials. The products when ready will be a huge demand for the people too.\n\nSupply:-“ Supply means the fulfilment of demand. In a large and highly populous country like India, there is always a demand for finished products. If more IT cities do not develop, the companies cannot fulfil the needs and desires of the people of a populous country like India. As IT cities develop, more IT companies will come, which will supply more and more finished IT products to our people.\n\nMarket: A market is a place where different economic agents like buyers and sellers interact with one another. In a populous country like India, there is a huge market. As IT cities will grow, more and more IT companies will come from across the world and more will the competition in the market increase. This will help consumers as they will get more and more differentiated products and the market will also run smoothly. A competitive market is always good and healthy. We can safely assume that our oligopoly market will surely tend to reach a perfectly competitive market by the year 2040.\n\nRevenue:- As the market increases, more revenue will be generated. Now at present, the IT revenue of India is 245 million dollars, 19 million dollars more than the financial year 2022. If IT cities grow, then more companies will invest which leads to an increase in the IT market which in turn generates more revenue in India. We can expect that the IT revenue of India will cross or nearly tend to reach 10 billion dollars by 2040.\n\nImpact on Environment\n\nThe rising IT cities will create a huge impact on the environment, the maximum of which will be harmful effects. The impact of rising IT cities on the environment is-\n\nDeforestation:- There will be cutting of trees in huge numbers to make the building of the IT companies which will cause great harm to the environment. The cutting of trees on a large scale will also cause mass degradation of forests.\n\nMore carbon footprint:- The IT companies will generate more carbon footprint in the atmosphere. South Asian countries including India are known for their lower carbon footprint. But if the IT sector grows this way then we will also be at the same pace of generation of carbon footprint by 2040.\n\nDeath of birds:- The

cell phone and mobile towers by the telecom companies caused the death of birds which caused a great imbalance in the ecosystem. The number of sparrows has been reduced due to this phenomenon. If this goes on we can see the extinction of many bird species by 2040.

**Impact on infrastructure**

There are many contributions of the IT cities on infrastructure. They are re-

**Transportation:-** The rising IT cities need an excellent transport system for the supply of raw materials and delivery of the finished products into the market. So the transportation system develops in that area. So we have an excellent transport system by 2040.

**Need for a public transport system:-** There is a need for a public transport system in the IT cities. As the IT cities are a source of employment and a huge population reside in these areas, there is an adequate need for public transport systems like buses, taxis etc. We hope that it will be improved by 2040.

**Water supply:-** As a huge number of people reside in the IT cities there is a need for a dequate water supply to fulfil the needs of people as well as for industries. This will help us to find many new methods of water supply and conservation by 2040.

**Electricity:-** Electric supply is the lifeline of the sector. Without an electric supply, no machines will run and not even the IT cities will flourish. If the IT cities flourish this way, we going to have an excellent electric supply by 2040.

**Healthcare:-** As a large number of people reside in IT cities, there is a need for proper health infrastructure and healthcare facilities for the people. So with the growth of IT cities, our healthcare system will also improve by 2040.

**Education:-** Education is the primary key or core of any nation. There must be proper education and training centres in those IT cities to fulfil the people's demands. So with the growth of IT cities, the education system will also develop by 2040. Our education is also going to be skill-oriented.

**Impact on city life**

With the growth of IT cities, more people will get jobs and will earn more. So the purchasing power of the people will increase. People will lead a better lifestyle. They will buy things of good brand value. The tastes and preferences of people will also change. The human development index is going to increase. People will buy good quality food and good quality cars. So the food, automobile and many other industries are going to increase. So there will be a huge impact on city life by 2040.

Blackcoffer Insights 47: Arka Mukhopadhyay, West Bengal University Of Animal And Fishery Sciences

```
In [22]: cleaned_text = " ".join(clean_text(input_data))
positive_score = calculate_positive_score(cleaned_text, positive_word_dic)
negative_score = calculate_negative_score(cleaned_text, negative_word_dic)
polarity_score = calculate_polarity_score(positive_score, negative_score)
subjectivity_score = calculate_subjectivity_score(positive_score, negative_score)

avg_sentence_length, percentage_complex_words, fog_index, avg_words_per_sentence,
syllables_per_word = calculate_syllables_per_word(input_data)
personal_pronouns = count_personal_pronouns(input_data)
avg_word_length = calculate_average_word_length(input_data)
```

```
In [23]: output_data.append([positive_score, negative_score, polarity_score, subjectivity_score,
                             avg_sentence_length, percentage_complex_words, fog_index,
                             avg_words_per_sentence, complex_word_count, word_count,
                             syllables_per_word, personal_pronouns, avg_word_length])
```

```
In [24]: output_columns = ["POSITIVE SCORE", "NEGATIVE SCORE", "POLARITY SCORE", "SU  
"AVG SENTENCE LENGTH", "PERCENTAGE OF COMPLEX WORDS", "FO  
"AVG NUMBER OF WORDS PER SENTENCE", "COMPLEX WORD COUNT",  
"SYLLABLE PER WORD", "PERSONAL PRONOUNS", "AVG WORD LENGT
```

```
In [25]: output_df = pd.DataFrame(output_data, columns=output_columns)  
  
# Save the output DataFrame to Excel  
output_df.to_excel("Output Data Structure.xlsx", index=False)
```

```
In [ ]:
```