Assigment7

```
pip install nltk
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: nltk in /usr/local/lib/python3.9/dist-packages (3.8.1)
Requirement already satisfied: tqdm in /usr/local/lib/python3.9/dist-packages (from nltk) (4.65.0)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.9/dist-packages (from nltk) (2022.10.31)
Requirement already satisfied: joblib in /usr/local/lib/python3.9/dist-packages (from nltk) (1.1.1)
Requirement already satisfied: click in /usr/local/lib/python3.9/dist-packages (from nltk) (8.1.3)
```

```
import nltk
```

```
nltk.__version__
```

```
'3.8.1'
```

```
from nltk.tokenize import sent_tokenize
```

```
text="""Hello Mr. Green, how are you doing today? The weather is great, and city is awesome. The sky is pinkish blue. You shouldnt eat ca
```

```
nltk.download('punkt')
```

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

```
type(text)
```

```
str
```

```
tokenized_sent= sent_tokenize(text)
```

```
tokenized_sent
```

```
['Hello Mr. Green, how are you doing today?',
 'The weather is great, and city is awesome.',
 'The sky is pinkish blue.',
 'You shouldnt eat cakes']
```

```
from nltk.tokenize import word_tokenize
tokenized_word= word_tokenize(text)
print(tokenized_word)
```

```
['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ',', 'and', 'city', 'is
```

```
len(tokenized_word)
```

```
30
```

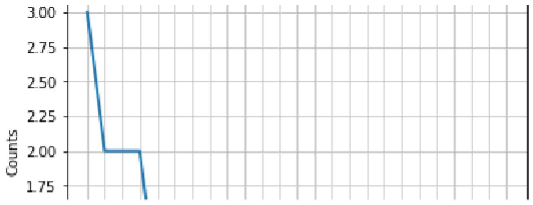
```
from nltk.probability import FreqDist
fdist= FreqDist(tokenized_word)
print(fdist)
```

```
<FreqDist with 25 samples and 30 outcomes>
```

```
fdist.most_common(3)
```

```
[('is', 3), (',', 2), ('The', 2)]
```

```
import matplotlib.pyplot as plt
fdist.plot(30,cumulative=False)
plt.show()
plt.figure(figsize=(12,12))
```




```
from nltk.corpus import stopwords
```

```
nltk.download('stopwords')
```

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

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

```
{'both', 'shan', "it's", 'into', 'above', 'about', 'yourselves', 'same', 'those', "shan't", 'having', 'hers', 'be', "weren't", 'how
```

```
len(stop_words)
```

```
179
```

```
filtered_sent=[]
for w in tokenized_sent:
  if w not in stop_words:
    filtered_sent.append(w)
    print("Tokenized Sentence:",tokenized_sent)
    print("Filtered Sentence:",filtered_sent)
```

```
Tokenized Sentence: ['Hello Mr. Green, how are you doing today?', 'The weather is great, and city is awesome.', 'The sky is pinkish
Filtered Sentence: ['Hello Mr. Green, how are you doing today?']
Tokenized Sentence: ['Hello Mr. Green, how are you doing today?', 'The weather is great, and city is awesome.', 'The sky is pinkish
Filtered Sentence: ['Hello Mr. Green, how are you doing today?', 'The weather is great, and city is awesome.']
Tokenized Sentence: ['Hello Mr. Green, how are you doing today?', 'The weather is great, and city is awesome.', 'The sky is pinkish
Filtered Sentence: ['Hello Mr. Green, how are you doing today?', 'The weather is great, and city is awesome.', 'The sky is pinkish
Tokenized Sentence: ['Hello Mr. Green, how are you doing today?', 'The weather is great, and city is awesome.', 'The sky is pinkish
Filtered Sentence: ['Hello Mr. Green, how are you doing today?', 'The weather is great, and city is awesome.', 'The sky is pinkish
```

```
filtered_sent=[]
for w in tokenized_word:
  if w not in stop_words:
    filtered_sent.append(w)
    print("Tokenized Sentence:",tokenized_word)
    print("Filtered Sentence:",filtered_sent)
```

```
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome', '.']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome', '.', 'The']
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome', '.', 'The', 'sky
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome', '.', 'The', 'sky
```

```
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome', '.', 'The', 'sky
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome', '.', 'The', 'sky
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome', '.', 'The', 'sky
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome', '.', 'The', 'sky
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome', '.', 'The', 'sky
Tokenized Sentence: ['Hello', 'Mr.', 'Green', ',', 'how', 'are', 'you', 'doing', 'today', '?', 'The', 'weather', 'is', 'great', ','
Filtered Sentence: ['Hello', 'Mr.', 'Green', ',', 'today', '?', 'The', 'weather', 'great', ',', 'city', 'awesome', '.', 'The', 'sky
```

```
len(tokenized_word)
```

```
30
```

```
len(filtered_sent)
```

```
22
```

```
from nltk.stem import PorterStemmer
```

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

```
ps=PorterStemmer()
```

```
stemmed_words=[]
for w in filtered_sent:
  stemmed_words.append(ps.stem(w))
  print("Filtered Sentence:",filtered_sent)
```

```
print("Stemmed Sentence:", stemmed_words)
```

```
Stemmed Sentence: ['hello', 'mr.', 'green', ',', 'today', '?', 'the', 'weather', 'great', ',', 'citi', 'awesom', '.', 'the', 'sky',
```

```
len(stemmed_words)
```

```
22
```

```
import nltk
```

```
nltk.download('wordnet')
```

```
[nltk_data] Downloading package wordnet to /root/nltk_data...
True
```

```
from nltk.stem.wordnet import WordNetLemmatizer
lem = WordNetLemmatizer()
```

```
from nltk.stem.porter import PorterStemmer
stem = PorterStemmer()


word = "flying"
print("Lemmatized Word:",lem.lemmatize(word,"v"))
print("Stemmed Word:",stem.stem(word))
```

```
Lemmatized Word: fly
Stemmed Word: fli
```

```
nltk.download('punkt')
```

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

```
sent = "Albert Einstein was born in Ulm, Germany in 1879."
tokens=nltk.word_tokenize(sent)
print(tokens)
```

```
['Albert', 'Einstein', 'was', 'born', 'in', 'Ulm', ',', 'Germany', 'in', '1879', '.']
```

```
nltk.download('averaged_perceptron_tagger')
```

```
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /root/nltk_data...
[nltk_data]   Unzipping taggers/averaged_perceptron_tagger.zip.
True
```

```
nltk.pos_tag(tokens)
```

```
[('Albert', 'NNP'),
 ('Einstein', 'NNP'),
 ('was', 'VBD'),
 ('born', 'VBN'),
 ('in', 'IN'),
 ('Ulm', 'NNP'),
 (',', ','),
 ('Germany', 'NNP'),
 ('in', 'IN'),
 ('1879', 'CD'),
 ('.', '.')]
```