1) Display the list of unigram , bigram and trigram possibilities for the chosen corpus text

Input –

```python
import nltk
from nltk.util import ngrams
import string

nltk.download('punkt')

def remove_punct(corpus_text):
    punct = string.punctuation
    filtered_sent = ""

    for str in corpus_text:
        if str not in punct:
            filtered_sent += str

    return filtered_sent

corpus_text = '''Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.'''

words = nltk.word_tokenize(remove_punct(corpus_text))

unigrams = list(ngrams(words, 1))
bigrams = list(ngrams(words, 2))
trigrams = list(ngrams(words, 3))

print("Unigrams:", unigrams)
print(" \n Bigrams:", bigrams)
print("\nTrigrams:", trigrams)
```

Output –

Unigrams: [('Two',), ('roads',), ('diverged',), ('in',), ('a',), ('yellow',), ('wood',), ('And',), ('sorry',), ('I',), ('could',), ('not',), ('travel',), ('both',), ('And',), ('be',), ('one',), ('traveler',), ('long',), ('I',), ('stood',), ('And',), ('looked',), ('down',), ('one',), ('as',), ('far',), ('as',), ('I',), ('could',), ('To',), ('where',), ('it',), ('bent',), ('in',), ('the',), ('undergrowth',), ('Then',), ('took',), ('the',), ('other',), ('as',), ('just',), ('as',), ('fair',), ('And',), ('having',), ('perhaps',), ('the',), ('better',), ('claim',), ('Because',), ('it',), ('was',), ('grassy',), ('and',), ('wanted',), ('wear',), ('Though',), ('as',), ('for',), ('that',), ('the',), ('passing',), ('there',), ('Had',), ('worn',), ('them',), ('really',), ('about',), ('the',), ('same',), ('And',), ('both',), ('that',), ('morning',), ('equally',), ('lay',), ('In',), ('leaves',), ('no',), ('step',), ('had',), ('trodden',), ('black',), ('Oh',), ('I',), ('kept',), ('the',), ('first',), ('for',), ('another',), ('day',), ('Yet',), ('knowing',), ('how',), ('way',), ('leads',), ('on',), ('to',), ('way',), ('I',), ('doubted',), ('if',), ('I',), ('should',), ('ever',), ('come',), ('back',), ('I',), ('shall',), ('be',), ('telling',), ('this',), ('with',), ('a',), ('sigh',), ('Somewhere',), ('ages',), ('and',), ('ages',), ('hence',), ('Two',), ('roads',), ('diverged',), ('in',), ('a',), ('wood',), ('and',), ('I-',), ('I',), ('took',), ('the',), ('one',), ('less',), ('traveled',), ('by',), ('And',), ('that',), ('has',), ('made',), ('all',), ('the',), ('difference',)]

Bigrams: [('Two', 'roads'), ('roads', 'diverged'), ('diverged', 'in'), ('in', 'a'), ('a', 'yellow'), ('yellow', 'wood'), ('wood', 'And'), ('And', 'sorry'), ('sorry', 'I'), ('I', 'could'), ('could', 'not'), ('not', 'travel'), ('travel', 'both'), ('both', 'And'), ('And', 'be'), ('be', 'one'), ('one', 'traveler'), ('traveler', 'long'), ('long', 'I'), ('I', 'stood'), ('stood', 'And'), ('And', 'looked'), ('looked', 'down'), ('down', 'one'), ('one', 'as'), ('as', 'far'), ('far', 'as'), ('as', 'I'), ('I', 'could'), ('could', 'To'), ('To', 'where'), ('where', 'it'), ('it', 'bent'), ('bent', 'in'), ('in', 'the'), ('the', 'undergrowth'), ('undergrowth', 'Then'), ('Then', 'took'), ('took', 'the'), ('the', 'other'), ('other', 'as'), ('as', 'just'), ('just', 'as'), ('as', 'fair'), ('fair', 'And'), ('And', 'having'), ('having', 'perhaps'), ('perhaps', 'the'), ('the', 'better'), ('better', 'claim'), ('claim', 'Because'), ('Because', 'it'), ('it', 'was'), ('was', 'grassy'), ('grassy', 'and'), ('and', 'wanted'), ('wanted', 'wear'), ('wear', 'Though'), ('Though', 'as'), ('as', 'for'), ('for', 'that'), ('that', 'the'), ('the', 'passing'), ('passing', 'there'), ('there', 'Had'), ('Had', 'worn'), ('worn', 'them'), ('them', 'really'), ('really', 'about'), ('about', 'the'), ('the', 'same'), ('same', 'And'), ('And', 'both'), ('both', 'that'), ('that', 'morning'), ('morning', 'equally'), ('equally', 'lay'), ('lay', 'In'), ('In', 'leaves'), ('leaves', 'no'), ('no', 'step'), ('step', 'had'), ('had', 'trodden'), ('trodden', 'black'), ('black', 'Oh'), ('Oh', 'I'), ('I', 'kept'), ('kept', 'the'), ('the', 'first'), ('first', 'for'), ('for', 'another'), ('another', 'day'), ('day', 'Yet'), ('Yet', 'knowing'), ('knowing', 'how'), ('how', 'way'), ('way', 'leads'), ('leads', 'on'), ('on', 'to'), ('to', 'way'), ('way', 'I'), ('I', 'doubted'), ('doubted', 'if'), ('if', 'I'), ('I', 'should'), ('should', 'ever'), ('ever', 'come'), ('come', 'back'), ('back', 'I'), ('I', 'shall'), ('shall', 'be'), ('be', 'telling'), ('telling', 'this'), ('this', 'with'), ('with', 'a'), ('a', 'sigh'), ('sigh', 'Somewhere'), ('Somewhere', 'ages'), ('ages', 'and'), ('and', 'ages'), ('ages', 'hence'), ('hence', 'Two'), ('Two', 'roads'), ('roads', 'diverged'), ('diverged', 'in'), ('in', 'a'), ('a', 'wood'), ('wood', 'and'), ('and', 'I-'), ('I-', 'I'), ('I', 'took'), ('took', 'the'), ('the', 'one'), ('one', 'less'), ('less', 'traveled'), ('traveled', 'by'), ('by', 'And'), ('And', 'that'), ('that', 'has'), ('has', 'made'), ('made', 'all'), ('all', 'the'), ('the', 'difference')]

Trigrams: [('Two', 'roads', 'diverged'), ('roads', 'diverged', 'in'), ('diverged', 'in', 'a'), ('in', 'a', 'yellow'), ('a', 'yellow', 'wood'), ('yellow', 'wood', 'And'), ('wood', 'And', 'sorry'), ('And', 'sorry', 'I'), ('sorry', 'I', 'could'), ('I', 'could', 'not'), ('could', 'not', 'travel'), ('not', 'travel', 'both'), ('travel', 'both', 'And'), ('both', 'And', 'be'), ('And', 'be', 'one'), ('be', 'one', 'traveler'), ('one', 'traveler', 'long'), ('traveler', 'long', 'I'), ('long', 'I', 'stood'), ('I', 'stood', 'And'), ('stood', 'And', 'looked'), ('And', 'looked', 'down'), ('looked', 'down', 'one'), ('down', 'one', 'as'), ('one', 'as', 'far'), ('as', 'far', 'as'), ('far', 'as', 'I'), ('as', 'I', 'could'), ('I', 'could', 'To'), ('could', 'To', 'where'), ('To', 'where', 'it'), ('where', 'it', 'bent'), ('it', 'bent', 'in'), ('bent', 'in', 'the'), ('in', 'the', 'undergrowth'), ('the', 'undergrowth', 'Then'), ('undergrowth', 'Then', 'took'), ('Then', 'took', 'the'), ('took', 'the', 'other'), ('the', 'other', 'as'), ('other', 'as', 'just'), ('as', 'just', 'as'), ('just', 'as', 'fair'), ('as', 'fair', 'And'), ('fair', 'And', 'having'), ('And', 'having', 'perhaps'), ('having', 'perhaps', 'the'), ('perhaps', 'the', 'better'), ('the', 'better', 'claim'), ('better', 'claim', 'Because'), ('claim', 'Because', 'it'), ('Because', 'it', 'was'), ('it', 'was', 'grassy'), ('was', 'grassy', 'and'), ('grassy', 'and', 'wanted'), ('and', 'wanted', 'wear'), ('wanted', 'wear', 'Though'), ('wear', 'Though', 'as'), ('Though', 'as', 'for'), ('as', 'for', 'that'), ('for', 'that', 'the'), ('that', 'the', 'passing'), ('the', 'passing', 'there'), ('passing', 'there', 'Had'), ('there', 'Had', 'worn'), ('Had', 'worn', 'them'), ('worn', 'them', 'really'), ('them', 'really', 'about'), ('really', 'about', 'the'), ('about', 'the', 'same'), ('the', 'same', 'And'), ('same', 'And', 'both'), ('And', 'both', 'that'), ('both', 'that', 'morning'), ('that', 'morning', 'equally'), ('morning', 'equally', 'lay'), ('equally', 'lay', 'In'), ('lay', 'In', 'leaves'), ('In', 'leaves', 'no'), ('leaves', 'no', 'step'), ('no', 'step', 'had'), ('step', 'had', 'trodden'), ('had', 'trodden', 'black'), ('trodden', 'black', 'Oh'), ('black', 'Oh', 'I'), ('Oh', 'I', 'kept'), ('I', 'kept', 'the'), ('kept', 'the', 'first'), ('the', 'first', 'for'), ('first', 'for', 'another'), ('for', 'another', 'day'), ('another', 'day', 'Yet'), ('day', 'Yet', 'knowing'), ('Yet', 'knowing', 'how'), ('knowing', 'how', 'way'), ('how', 'way', 'leads'), ('way', 'leads', 'on'), ('leads', 'on', 'to'), ('on', 'to', 'way'), ('to', 'way', 'I'), ('way', 'I', 'doubted'), ('I', 'doubted', 'if'), ('doubted', 'if', 'I'), ('if', 'I', 'should'), ('I', 'should', 'ever'), ('should', 'ever', 'come'), ('ever', 'come', 'back'), ('come', 'back', 'I'), ('back', 'I', 'shall'), ('I', 'shall', 'be'), ('shall', 'be', 'telling'), ('be', 'telling', 'this'), ('telling', 'this', 'with'), ('this', 'with', 'a'), ('with', 'a', 'sigh'), ('a', 'sigh', 'Somewhere'), ('sigh', 'Somewhere', 'ages'), ('Somewhere', 'ages', 'and'), ('ages', 'and', 'ages'), ('and', 'ages', 'hence'), ('ages', 'hence', 'Two'), ('hence', 'Two', 'roads'), ('Two', 'roads', 'diverged'), ('roads', 'diverged', 'in'), ('diverged', 'in', 'a'), ('in', 'a', 'wood'), ('a', 'wood', 'and'), ('wood', 'and', 'I-'), ('and', 'I-', 'I'), ('I-', 'I', 'took'), ('I', 'took', 'the'), ('took', 'the', 'one'), ('the', 'one', 'less'), ('one', 'less', 'traveled'), ('less', 'traveled', 'by'), ('traveled', 'by', 'And'), ('by', 'And', 'that'), ('And', 'that', 'has'), ('that', 'has', 'made'), ('has', 'made', 'all'), ('made', 'all', 'the'), ('all', 'the', 'difference')]

2) Display the bigram probability table 1st table bigram counts 2nd table bigram probabilities

Input –

```python
import nltk
from nltk.util import bigrams
from nltk.probability import FreqDist, ConditionalFreqDist

nltk.download('punkt')

def remove_punct(corpus_text):
    punct = string.punctuation
    filtered_sent = ""

    for str in corpus_text:
        if str not in punct:
            filtered_sent += str

    return filtered_sent

corpus_text = '''Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;
Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,
And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.
I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.'''

words = nltk.word_tokenize(remove_punct(corpus_text))
```

```
bi_grams = list(bigrams(words))

bigram_freq = FreqDist(bi_grams) # It's a dictonary that stores how many times a bigram occured - ('Two', 'roads') : 2

bigram_cond_freq = ConditionalFreqDist(bi_grams) #Object

# Calculate bigram probabilities
bigram_probabilities = {}
for word1 in bigram_cond_freq:
    total_count = sum(bigram_cond_freq[word1].values())
    bigram_probabilities[word1] = {word2: count / total_count for word2, count in bigram_cond_freq[word1].items()}

# Display bigram counts
print("Bigram Counts:")
for bigram, count in bigram_freq.items():
    print(bigram, ":", count)

# Display bigram probabilities
print("\nBigram Probabilities:")
for word1, word2_prob in bigram_probabilities.items():
    for word2, prob in word2_prob.items():
        print(f"{word1} -> {word2} : {prob:.4f}")
```

Output –

```
Bigram Counts:
('Two', 'roads') : 2
('roads', 'diverged') : 2
('diverged', 'in') : 2
('in', 'a') : 2
('a', 'yellow') : 1
('yellow', 'wood') : 1
('wood', 'And') : 1
('And', 'sorry') : 1
('sorry', 'I') : 1
('I', 'could') : 2
('could', 'not') : 1
('not', 'travel') : 1
('travel', 'both') : 1
('both', 'And') : 1
('And', 'be') : 1
('be', 'one') : 1
('one', 'traveler') : 1
('traveler', 'long') : 1

Bigram Probabilities:
Two -> roads : 1.0000
roads -> diverged : 1.0000
diverged -> in : 1.0000
in -> a : 0.6667
in -> the : 0.3333
a -> yellow : 0.3333
a -> sigh : 0.3333
a -> wood : 0.3333
yellow -> wood : 1.0000
wood -> And : 0.5000
wood -> and : 0.5000
And -> sorry : 0.1667
And -> be : 0.1667
And -> looked : 0.1667
And -> having : 0.1667
And -> both : 0.1667
And -> that : 0.1667
```

Repo Link

https://github.com/Shreyaww/Sem7_NLP/blob/main/Experiment_5_N-gram.ipynb