

1) Write a program (using nltk toolkit in python environment) to tokenize

a) A sentence

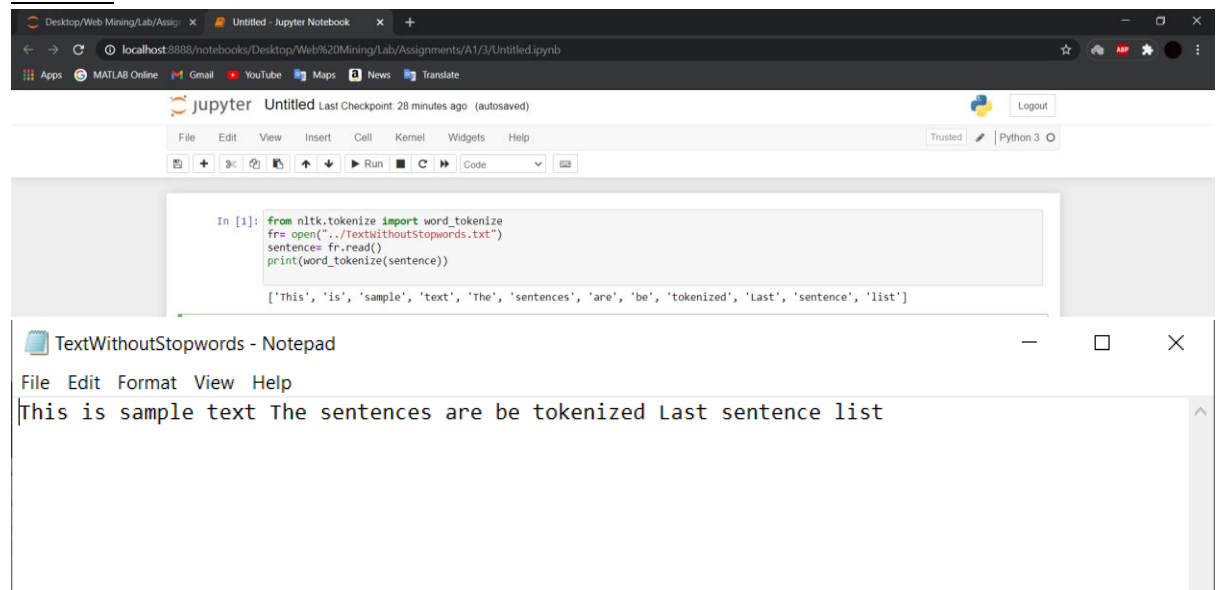
PROCEDURE

```
//Importing NLTK library to use the word_tokenize function
Import nltk.tokenize.word_tokenize
FileReader=open(File, read)
Sentence=FileReader.read()
Print(word_tokenize(Sentence))
```

CODE

```
from nltk.tokenize import word_tokenize
fr= open("../TextWithoutStopwords.txt")
sentence= fr.read()
print(word_tokenize(sentence))
```

OUTPUT



b) A paragraph

PROCEDURE

```
//Importing NLTK library to use the sent_tokenize function
Import nltk.tokenize.sent_tokenize
```

Registration Number:- 19BCE2119
Course:- Web Mining (L5+L6)

Name:- Gaurav Kumar Singh

```
FileReader=open(File, read)  
Paragraph=FileReader.read()  
Print(sent_tokenize(Paragraph))
```

CODE

```
from nltk.tokenize import sent_tokenize  
fr= open("../SampleText.txt")  
para= fr.read()  
print(sent_tokenize(para))
```

OUTPUT

