```
#!/usr/bin/env python

# coding: utf-8


# # Conversion lowercase


# In[1]:




import nltk

import string

import re



# In[20]:




def text_lowercase(text):

    return text.lower()


input_str = "Hey, did you know that the summer break is coming? Amazing right !! It's only 5 more days !!"

text_lowercase(input_str)




# # Number remove
```

```
# In[22]:



def remove_numbers(text):
    result = re.sub(r'\d+', '', text)
    return result

input_str = "There are 3 balls in this bag, and 12 in the other one."
remove_numbers(input_str)



# # Number to Text using inflect lib

# In[19]:



import inflect
p = inflect.engine()

# convert number into words
def convert_number(text):
    # split string into list of words
    temp_str = text.split()
    # initialise empty list
    new_string = []
```

```
    for word in temp_str:
        # if word is a digit, convert the digit
        # to numbers and append into the new_string list
        if word.isdigit():
            temp = p.number_to_words(word)
            new_string.append(temp)

        # append the word as it is
        else:
            new_string.append(word)

    # join the words of new_string to form a string
    temp_str = ' '.join(new_string)
    return temp_str

input_str = 'There are 3 balls in this bag, and 12 in the other one.'
convert_number(input_str)



# # OUTPUT : There are three balls in this bag, and twelve in the other one.


# # remove punctuation


# In[23]:
```

```
def remove_punctuation(text):
    translator = str.maketrans('', '', string.punctuation)
    return text.translate(translator)

input_str = "Hey, did you know that the summer break is coming? Amazing right !! It's only 5 more days !!"
remove_punctuation(input_str)



# # remove whitespace from text

# In[6]:


def remove_whitespace(text):
    return  " ".join(text.split())

input_str = "   we don't need   the given questions"
remove_whitespace(input_str)



# # Removal of Stopwords

# In[10]:
```

```
nltk.download('stopwords')



# In[18]:



from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
# remove stopwords function
def remove_stopwords(text):
    stop_words = set(stopwords.words("english"))
    word_tokens = word_tokenize(text)
    filtered_text = [word for word in word_tokens if word not in stop_words]
    return filtered_text

example_text = "This is a sample sentence and we are going to remove the stopwords from this."
remove_stopwords(example_text)



# In[ ]:
```

In[12]:

```
from nltk.stem.porter import PorterStemmer
from nltk.tokenize import word_tokenize
stemmer = PorterStemmer()

# stem words in the list of tokenized words
def stem_words(text):
    word_tokens = word_tokenize(text)
    stems = [stemmer.stem(word) for word in word_tokens]
    return stems

text = 'data science uses scientific methods algorithms and many types of processes'
stem_words(text)
```

In[ ]: