import pandas as pd

import nltk

from nltk.tokenize import word\_tokenize

from nltk.corpus import stopwords

from sklearn.feature\_extraction.text import TfidfVectorizer

pd.set\_option('max\_colwidth', 100)

nltk.download('punkt')

document\_df = pd.read\_csv("nlptext.csv", encoding='ISO-8859–1')

document1 = str(document\_df['text'])

document1 = "A paragraph is a series of related sentences developing a central idea, called the topic. Try to think about paragraphs in terms of thematic unity: a paragraph is a sentence or a group of sentences that supports one central, unified idea. Paragraphs add one idea at a time to your broader argument."

document1 = document1.lower()

words = word\_tokenize(document1)

print(words)

nltk.download('stopwords')

stopwords.words("english")

words\_new = []

for i in words:

if i not in stopwords.words("english"):

words\_new.append(i)

words\_new

words\_new.count('python')

for i in words\_new:

print(i , words\_new.count(i))

mydict = {}

for i in words:

mydict[i] = words\_new.count(i)

print(mydict)

d = {'one':1,'three':3,'five':5,'two':2,'four':4}

a = sorted(mydict.items(), key=lambda x: x[1], reverse=True)

print(a)