

# BAN200 Week 02 Homework

To complete the homework you will need to modify this template by adding Python code and/or text.

Before starting the homework, make sure to save a copy of this template to your personal Google Drive. If you haven't saved your own copy, any changes you make will be lost when you close your browser window.

To submit your homework: go to "File" in the Colab menu bar > select "Download" > select "Download .ipynb". This will download a ".ipynb" file to your computer. You must submit this file.

The homework is to be completed in groups. It is due at the start of next class.

Homework is graded on the following scale:

- 100% -- The assignment was submitted on time, any code runs without errors, and every question is answered correctly.
- 80% -- The assignment was submitted on time, any code runs without errors, and every question is answered. Some questions may be incorrect, but the submission demonstrates an average level of effort and average level of understanding of the material.
- 60% -- The submission demonstrates a below-average level of effort and below-average level of understanding of the material. This is the highest grade that should be given to submissions that are submitted late, have code that throws uncaught errors, or leave some questions unanswered.
- 0% -- No assignment was submitted, or the submission demonstrates little-to-no effort and little-to-no understanding of the material.

## Question 1

Create a function that takes a string and returns a list of word tokens, ignoring any non-alphanumeric characters. You should use `re.findall` in your function. Test your function with the string "Hello there, how are you?" and store the tokens in a variable `tokens`.

```
In [4]: # put your answer here
import re

tokens= "Hello there, how are you?"
```

## Question 2

Create a function that takes a list of tokens and makes them lowercase. Tip: you did this already in the Week 1 Homework. Test it with your `tokens` variable.

```
In [5]: # put your answer here
tokens=tokens.split()
for term in tokens:
```

hello  
there,  
how  
are  
you?

## Stop Words and Lemmas

The following code downloads stop words and lemmas and stores them in the variables `stopwords` and `lemmas` .

```
In [15]: import urllib.request, json
with urllib.request.urlopen("https://storage.googleapis.com/wd13/stopwords%20and%20lemma
    data = json.load(url)
    stopwords = data['stopwords']
    lemmas = data['lemmas']
```

`stopwords` is a list of stop words.

```
In [16]: stopwords
```

```
Out[16]: ['i',  
          'me',  
          'my',  
          'myself',  
          'we',  
          'our',  
          'ours',  
          'ourselves',  
          'you',  
          "you're",  
          "you've",  
          "you'll",  
          "you'd",  
          'your',  
          'yours',  
          'yourself',  
          'yourselves',  
          'he',  
          'him',  
          'his',  
          'himself',  
          'she',  
          "she's",  
          'her',  
          'hers',  
          'herself',  
          'it',  
          "it's",  
          'its',  
          'itself',  
          'they',  
          'them',  
          'their',  
          'theirs',  
          'themselves',  
          'what',  
          'which',  
          'who',  
          'whom',  
          'this',  
          'that',  
          "that'll",  
          'these',  
          'those',  
          'am',  
          'is',  
          'are',  
          'was',  
          'were',  
          'be',  
          'been',  
          'being',  
          'have',  
          'has',  
          'had',  
          'having',  
          'do',  
          'does',  
          'did',  
          'doing',  
          'a',  
          'an',  
          'the',  
          'and']
```

'but',  
'if',  
'or',  
'because',  
'as',  
'until',  
'while',  
'of',  
'at',  
'by',  
'for',  
'with',  
'about',  
'against',  
'between',  
'into',  
'through',  
'during',  
'before',  
'after',  
'above',  
'below',  
'to',  
'from',  
'up',  
'down',  
'in',  
'out',  
'on',  
'off',  
'over',  
'under',  
'again',  
'further',  
'then',  
'once',  
'here',  
'there',  
'when',  
'where',  
'why',  
'how',  
'all',  
'any',  
'both',  
'each',  
'few',  
'more',  
'most',  
'other',  
'some',  
'such',  
'no',  
'nor',  
'not',  
'only',  
'own',  
'same',  
'so',  
'than',  
'too',  
'very',  
's',  
't'.

```
'can',  
'will',  
'just',  
'don',  
"don't",  
'should',  
"should've",  
'now',  
'd',  
'll',  
'm',  
'o',  
're',  
've',  
'y',  
'ain',  
'aren',  
"aren't",  
'couldn',  
"couldn't",  
'didn',  
"didn't",  
'doesn',  
"doesn't",  
'hadn',  
"hadn't",  
'hasn',  
"hasn't",  
'haven',  
"haven't",  
'isn',  
"isn't",  
'ma',  
'mightn',  
"mightn't",  
'mustn',  
"mustn't",  
'needn',  
"needn't",  
'shan',  
"shan't",  
'shouldn',  
"shouldn't",  
'wasn',  
"wasn't",  
'weren',  
"weren't",  
'won',  
"won't",  
'wouldn',  
"wouldn't"]
```

`lemmas` is a dictionary where the "key" is a word and the "value" is the word's lemma.

In [7]: `lemmas`

```
Out[7]: {'aah': 'aah',
'aahed': 'aah',
'aahhing': 'aah',
'aahing': 'aah',
'aahs': 'aah',
'abandon': 'abandon',
'abandoned': 'abandon',
'abandoning': 'abandon',
'abandons': 'abandon',
'abase': 'abase',
'abased': 'abase',
'abases': 'abase',
'abasing': 'abase',
'abate': 'abate',
'abated': 'abate',
'abates': 'abate',
'abating': 'abate',
'abattoir': 'abattoir',
'abattoirs': 'abattoir',
'abbey': 'abbey',
'abbeys': 'abbey',
'abbot': 'abbot',
'abbots': 'abbot',
'abbreviate': 'abbreviate',
'abbreviated': 'abbreviate',
'abbreviating': 'abbreviate',
'abbreviation': 'abbreviation',
'abbreviations': 'abbreviation',
'abdicate': 'abdicate',
'abdicated': 'abdicate',
'abdicates': 'abdicate',
'abdication': 'abdicate',
'abdominal': 'abdominal',
'abdominals': 'abdominal',
'abduct': 'abduct',
'abducted': 'abduct',
'abducting': 'abduct',
'abducts': 'abduct',
'abduction': 'abduction',
'abductions': 'abduction',
'abductor': 'abductor',
'abductors': 'abductor',
'aberration': 'aberration',
'aberrations': 'aberration',
'abet': 'abet',
'abets': 'abet',
'abetted': 'abet',
'abetting': 'abet',
'abhor': 'abhor',
'abhorred': 'abhor',
'abhorring': 'abhor',
'abhors': 'abhor',
'abide': 'abide',
'abided': 'abide',
'abides': 'abide',
'abiding': 'abide',
'ability': 'ability',
'abilities': 'ability',
'abjure': 'abjure',
'abjured': 'abjure',
'abjures': 'abjure',
'abjuring': 'abjure',
'abl': 'abl',
'abler': 'abl'.
```

'ablest': 'abl',  
'ablution': 'ablution',  
'ablutions': 'ablution',  
'abnormality': 'abnormality',  
'abnormalities': 'abnormality',  
'abode': 'abode',  
'abodes': 'abode',  
'abolish': 'abolish',  
'abolished': 'abolish',  
'abolishes': 'abolish',  
'abolishing': 'abolish',  
'abolitionist': 'abolitionist',  
'abolitionists': 'abolitionist',  
'abomination': 'abomination',  
'abominations': 'abomination',  
'aborigine': 'aborigine',  
'aborigines': 'aborigine',  
'abort': 'abort',  
'aborted': 'abort',  
'aborting': 'abort',  
'aborts': 'abort',  
'abortion': 'abortion',  
'abortions': 'abortion',  
'abound': 'abound',  
'abounded': 'abound',  
'abounding': 'abound',  
'abounds': 'abound',  
'abrade': 'abrade',  
'abraded': 'abrade',  
'abrading': 'abrade',  
'abram': 'abram',  
'abrams': 'abram',  
'abrasion': 'abrasion',  
'abrasions': 'abrasion',  
'abrasive': 'abrasive',  
'abrasives': 'abrasive',  
'abridge': 'abridge',  
'abridged': 'abridge',  
'abridging': 'abridge',  
'abrogate': 'abrogate',  
'abrogated': 'abrogate',  
'abrogates': 'abrogate',  
'abrogating': 'abrogate',  
'abscess': 'abscess',  
'abscesses': 'abscess',  
'abscond': 'abscond',  
'absconded': 'abscond',  
'absconding': 'abscond',  
'absconds': 'abscond',  
'abseil': 'abseil',  
'abseiled': 'abseil',  
'abseiling': 'abseil',  
'abseilling': 'abseil',  
'abseils': 'abseil',  
'absence': 'absence',  
'absences': 'absence',  
'absent': 'absent',  
'absented': 'absent',  
'absenting': 'absent',  
'absents': 'absent',  
'absentee': 'absentee',  
'absentees': 'absentee',  
'absolute': 'absolute',  
'absolutes': 'absolute'.

'absolutest': 'absolute',  
 'absolve': 'absolve',  
 'absolved': 'absolve',  
 'absolves': 'absolve',  
 'absolving': 'absolve',  
 'absorb': 'absorb',  
 'absorbed': 'absorb',  
 'absorbing': 'absorb',  
 'absorbs': 'absorb',  
 'absorber': 'absorber',  
 'absorbers': 'absorber',  
 'abstain': 'abstain',  
 'abstained': 'abstain',  
 'abstaining': 'abstain',  
 'abstains': 'abstain',  
 'abstainer': 'abstainer',  
 'abstainers': 'abstainer',  
 'abstention': 'abstention',  
 'abstentions': 'abstention',  
 'abstract': 'abstract',  
 'abstracted': 'abstract',  
 'abstracting': 'abstract',  
 'abstracts': 'abstract',  
 'abstraction': 'abstraction',  
 'abstractions': 'abstraction',  
 'absurdity': 'absurdity',  
 'absurdities': 'absurdity',  
 'abundance': 'abundance',  
 'abundances': 'abundance',  
 'abuse': 'abuse',  
 'abused': 'abuse',  
 'abuses': 'abuse',  
 'abusing': 'abuse',  
 'abuser': 'abuser',  
 'abusers': 'abuser',  
 'abut': 'abut',  
 'abuts': 'abut',  
 'abutted': 'abut',  
 'abutting': 'abut',  
 'abutment': 'abutment',  
 'abutments': 'abutment',  
 'abyssinian': 'abyssinian',  
 'abyssinians': 'abyssinian',  
 'ac': 'ac',  
 'acs': 'ac',  
 'academic': 'academic',  
 'academics': 'academic',  
 'academical': 'academical',  
 'academicals': 'academical',  
 'academician': 'academician',  
 'academicians': 'academician',  
 'academy': 'academy',  
 'academies': 'academy',  
 'accede': 'accede',  
 'acceded': 'accede',  
 'accedes': 'accede',  
 'acceding': 'accede',  
 'accelerate': 'accelerate',  
 'accelerated': 'accelerate',  
 'accelerates': 'accelerate',  
 'accelerating': 'accelerate',  
 'acceleration': 'acceleration',  
 'accelerations': 'acceleration',  
 'accelerator': 'accelerator'.



'accelerators': 'accelerator',  
 'accent': 'accent',  
 'accented': 'accent',  
 'accenting': 'accent',  
 'accents': 'accent',  
 'accentuate': 'accentuate',  
 'accentuated': 'accentuate',  
 'accentuates': 'accentuate',  
 'accentuating': 'accentuate',  
 'accept': 'accept',  
 'accepted': 'accept',  
 'accepting': 'accept',  
 'accepts': 'accept',  
 'acceptance': 'acceptance',  
 'acceptances': 'acceptance',  
 'access': 'access',  
 'accessed': 'access',  
 'accesses': 'access',  
 'accessing': 'access',  
 'accession': 'accession',  
 'accessioned': 'accession',  
 'accessioning': 'accession',  
 'accessions': 'accession',  
 'accessory': 'accessory',  
 'accessories': 'accessory',  
 'accident': 'accident',  
 'accidents': 'accident',  
 'accidental': 'accidental',  
 'accidentals': 'accidental',  
 'acclaim': 'acclaim',  
 'acclaimed': 'acclaim',  
 'acclaiming': 'acclaim',  
 'acclaims': 'acclaim',  
 'acclamation': 'acclamation',  
 'acclamations': 'acclamation',  
 'acclimatise': 'acclimatise',  
 'acclimatised': 'acclimatise',  
 'acclimatising': 'acclimatise',  
 'acclimatize': 'acclimatize',  
 'acclimatized': 'acclimatize',  
 'acclimatizing': 'acclimatize',  
 'accolade': 'accolade',  
 'accolades': 'accolade',  
 'accommodate': 'accommodate',  
 'accommodated': 'accommodate',  
 'accommodates': 'accommodate',  
 'accommodating': 'accommodate',  
 'accommodation': 'accommodation',  
 'accommodations': 'accommodation',  
 'accompaniment': 'accompaniment',  
 'accompaniments': 'accompaniment',  
 'accompany': 'accompany',  
 'accompanied': 'accompany',  
 'accompanies': 'accompany',  
 'accompanying': 'accompany',  
 'accomplice': 'accomplice',  
 'accomplices': 'accomplice',  
 'accomplish': 'accomplish',  
 'accomplished': 'accomplish',  
 'accomplishes': 'accomplish',  
 'accomplishing': 'accomplish',  
 'accomplishment': 'accomplishment',  
 'accomplishments': 'accomplishment',  
 'accord': 'accord'.

'accorded': 'accord',  
 'according': 'accord',  
 'accords': 'accord',  
 'accordion': 'accordion',  
 'accordions': 'accordion',  
 'accost': 'accost',  
 'accosted': 'accost',  
 'accosting': 'accost',  
 'accosts': 'accost',  
 'account': 'account',  
 'accounted': 'account',  
 'accounting': 'account',  
 'accounts': 'account',  
 'accountability': 'accountability',  
 'accountabilities': 'accountability',  
 'accountant': 'accountant',  
 'accountants': 'accountant',  
 'accoutrement': 'accoutrement',  
 'accoutrements': 'accoutrement',  
 'accredit': 'accredit',  
 'accredited': 'accredit',  
 'accrediting': 'accredit',  
 'accredits': 'accredit',  
 'accrete': 'accrete',  
 'accreted': 'accrete',  
 'accretes': 'accrete',  
 'accreting': 'accrete',  
 'accretion': 'accretion',  
 'accretions': 'accretion',  
 'accrual': 'accrual',  
 'accruals': 'accrual',  
 'accrue': 'accrue',  
 'accrued': 'accrue',  
 'accrues': 'accrue',  
 'accruing': 'accrue',  
 'accumulate': 'accumulate',  
 'accumulated': 'accumulate',  
 'accumulates': 'accumulate',  
 'accumulating': 'accumulate',  
 'accumulation': 'accumulation',  
 'accumulations': 'accumulation',  
 'accumulator': 'accumulator',  
 'accumulators': 'accumulator',  
 'accusation': 'accusation',  
 'accusations': 'accusation',  
 'accuse': 'accuse',  
 'accused': 'accuse',  
 'accuses': 'accuse',  
 'accusing': 'accuse',  
 'accuser': 'accuser',  
 'accusers': 'accuser',  
 'accustom': 'accustom',  
 'accustomed': 'accustom',  
 'accustoms': 'accustom',  
 'ace': 'ace',  
 'aced': 'ace',  
 'aces': 'ace',  
 'acing': 'ace',  
 'acer': 'acer',  
 'acers': 'acer',  
 'acetylate': 'acetylate',  
 'acetylated': 'acetylate',  
 'acetylating': 'acetylate',  
 'ach': 'ach'

'aches': 'ach',  
 'ache': 'ache',  
 'ached': 'ache',  
 'aching': 'ache',  
 'achieve': 'achieve',  
 'achieved': 'achieve',  
 'achieveing': 'achieve',  
 'achieves': 'achieve',  
 'achieving': 'achieve',  
 'achievement': 'achievement',  
 'achievements': 'achievement',  
 'achiever': 'achiever',  
 'achievers': 'achiever',  
 'acid': 'acid',  
 'acids': 'acid',  
 'acidify': 'acidify',  
 'acidified': 'acidify',  
 'acidifies': 'acidify',  
 'acidifying': 'acidify',  
 'acknowledge': 'acknowledge',  
 'acknowledged': 'acknowledge',  
 'acknowledges': 'acknowledge',  
 'acknowledging': 'acknowledge',  
 'acknowledgement': 'acknowledgement',  
 'acknowledgements': 'acknowledgement',  
 'acknowledgment': 'acknowledgment',  
 'acknowledgments': 'acknowledgment',  
 'acolyte': 'acolyte',  
 'acolytes': 'acolyte',  
 'acorn': 'acorn',  
 'acorns': 'acorn',  
 'acoustic': 'acoustic',  
 'acoustics': 'acoustic',  
 'acquaint': 'acquaint',  
 'acquainted': 'acquaint',  
 'acquainting': 'acquaint',  
 'acquaints': 'acquaint',  
 'acquaintance': 'acquaintance',  
 'acquaintances': 'acquaintance',  
 'acquiesce': 'acquiesce',  
 'acquiesced': 'acquiesce',  
 'acquiesces': 'acquiesce',  
 'acquiescing': 'acquiesce',  
 'acquire': 'acquire',  
 'acquired': 'acquire',  
 'acquires': 'acquire',  
 'acquiring': 'acquire',  
 'acquirer': 'acquirer',  
 'acquirers': 'acquirer',  
 'acquisition': 'acquisition',  
 'acquisitions': 'acquisition',  
 'acquit': 'acquit',  
 'acquits': 'acquit',  
 'acquitted': 'acquit',  
 'acquitting': 'acquit',  
 'acquittal': 'acquittal',  
 'acquittals': 'acquittal',  
 'acre': 'acre',  
 'acres': 'acre',  
 'acrobat': 'acrobat',  
 'acrobats': 'acrobat',  
 'acrobatic': 'acrobatic',  
 'acrobatics': 'acrobatic',  
 'acronvm': 'acronvm'.

'acronymed': 'acronym',  
'acronyms': 'acronym',  
'acrylic': 'acrylic',  
'acrylics': 'acrylic',  
'act': 'act',  
'acted': 'act',  
'acting': 'act',  
'acts': 'act',  
'action': 'action',  
'actioned': 'action',  
'actioning': 'action',  
'actions': 'action',  
'activate': 'activate',  
'activated': 'activate',  
'activates': 'activate',  
'activating': 'activate',  
'activator': 'activator',  
'activators': 'activator',  
'active': 'active',  
'actives': 'active',  
'activist': 'activist',  
'activists': 'activist',  
'activity': 'activity',  
'activities': 'activity',  
'actor': 'actor',  
'actors': 'actor',  
'actress': 'actress',  
'actresses': 'actress',  
'actual': 'actual',  
'actuals': 'actual',  
'actuality': 'actuality',  
'actualities': 'actuality',  
'actualize': 'actualize',  
'actualized': 'actualize',  
'actualizes': 'actualize',  
'actualizing': 'actualize',  
'actuary': 'actuary',  
'actuaries': 'actuary',  
'actuate': 'actuate',  
'actuated': 'actuate',  
'actuates': 'actuate',  
'actuating': 'actuate',  
'acupuncturist': 'acupuncturist',  
'acupuncturists': 'acupuncturist',  
'ad': 'ad',  
'ads': 'ad',  
'adapt': 'adapt',  
'adapted': 'adapt',  
'adapting': 'adapt',  
'adapts': 'adapt',  
'adaptation': 'adaptation',  
'adaptations': 'adaptation',  
'adapter': 'adapter',  
'adapters': 'adapter',  
'adaptor': 'adaptor',  
'adaptors': 'adaptor',  
'add': 'add',  
'added': 'add',  
'adding': 'add',  
'adds': 'add',  
'adder': 'adder',  
'adders': 'adder',  
'addict': 'addict',  
'addicted': 'addict'.

'addicts': 'addict',  
 'addiction': 'addiction',  
 'addictions': 'addiction',  
 'addition': 'addition',  
 'additions': 'addition',  
 'additive': 'additive',  
 'additives': 'additive',  
 'addle': 'addle',  
 'addled': 'addle',  
 'addles': 'addle',  
 'address': 'address',  
 'addressed': 'address',  
 'addresses': 'address',  
 'addressing': 'address',  
 'addressee': 'addressee',  
 'addressees': 'addressee',  
 'adduce': 'adduce',  
 'adduced': 'adduce',  
 'adduces': 'adduce',  
 'adducing': 'adduce',  
 'adduct': 'adduct',  
 'adducted': 'adduct',  
 'adducts': 'adduct',  
 'adenocarcinoma': 'adenocarcinoma',  
 'adenocarcinomas': 'adenocarcinoma',  
 'adenoid': 'adenoid',  
 'adenoids': 'adenoid',  
 'adenoma': 'adenoma',  
 'adenomas': 'adenoma',  
 'adept': 'adept',  
 'adepts': 'adept',  
 'adhere': 'adhere',  
 'adhered': 'adhere',  
 'adhereing': 'adhere',  
 'adheres': 'adhere',  
 'adhering': 'adhere',  
 'adherent': 'adherent',  
 'adherents': 'adherent',  
 'adhesion': 'adhesion',  
 'adhesions': 'adhesion',  
 'adhesive': 'adhesive',  
 'adhesives': 'adhesive',  
 'adida': 'adida',  
 'adidas': 'adida',  
 'adio': 'adio',  
 'adios': 'adio',  
 'adit': 'adit',  
 'adits': 'adit',  
 'adjective': 'adjective',  
 'adjectives': 'adjective',  
 'adjoin': 'adjoin',  
 'adjoined': 'adjoin',  
 'adjoining': 'adjoin',  
 'adjoins': 'adjoin',  
 'adjourn': 'adjourn',  
 'adjourned': 'adjourn',  
 'adjourning': 'adjourn',  
 'adjourns': 'adjourn',  
 'adjudge': 'adjudge',  
 'adjudged': 'adjudge',  
 'adjudges': 'adjudge',  
 'adjudging': 'adjudge',  
 'adjudicate': 'adjudicate',  
 'adiudicated': 'adiudicate'.

'adjudicates': 'adjudicate',  
 'adjudicating': 'adjudicate',  
 'adjudication': 'adjudication',  
 'adjudications': 'adjudication',  
 'adjudicator': 'adjudicator',  
 'adjudicators': 'adjudicator',  
 'adjunct': 'adjunct',  
 'adjuncts': 'adjunct',  
 'adjure': 'adjure',  
 'adjured': 'adjure',  
 'adjuring': 'adjure',  
 'adjust': 'adjust',  
 'adjusted': 'adjust',  
 'adjusting': 'adjust',  
 'adjusts': 'adjust',  
 'adjuster': 'adjuster',  
 'adjusters': 'adjuster',  
 'adjustment': 'adjustment',  
 'adjustments': 'adjustment',  
 'adman': 'adman',  
 'admen': 'adman',  
 'administer': 'administer',  
 'administered': 'administer',  
 'administering': 'administer',  
 'administers': 'administer',  
 'administrate': 'administrate',  
 'administrated': 'administrate',  
 'administrates': 'administrate',  
 'administrating': 'administrate',  
 'administration': 'administration',  
 'administrations': 'administration',  
 'administrator': 'administrator',  
 'administrators': 'administrator',  
 'admiral': 'admiral',  
 'admirals': 'admiral',  
 'admire': 'admire',  
 'admired': 'admire',  
 'admires': 'admire',  
 'admiring': 'admire',  
 'admirer': 'admirer',  
 'admirers': 'admirer',  
 'admission': 'admission',  
 'admissions': 'admission',  
 'admit': 'admit',  
 'admitted': 'admit',  
 'admits': 'admit',  
 'admitted': 'admit',  
 'admitting': 'admit',  
 'admonish': 'admonish',  
 'admonished': 'admonish',  
 'admonishes': 'admonish',  
 'admonishing': 'admonish',  
 'admonition': 'admonition',  
 'admonitions': 'admonition',  
 'adolescent': 'adolescent',  
 'adolescents': 'adolescent',  
 'adopt': 'adopt',  
 'adopted': 'adopt',  
 'adopting': 'adopt',  
 'adopts': 'adopt',  
 'adoptee': 'adoptee',  
 'adoptees': 'adoptee',  
 'adopter': 'adopter',  
 'adonaters': 'adonater'.

'adoption': 'adoption',  
 'adoptions': 'adoption',  
 'adore': 'adore',  
 'adored': 'adore',  
 'adores': 'adore',  
 'adoring': 'adore',  
 'adorn': 'adorn',  
 'adorned': 'adorn',  
 'adorning': 'adorn',  
 'adorns': 'adorn',  
 'adornment': 'adornment',  
 'adornments': 'adornment',  
 'adsorb': 'adsorb',  
 'adsorbed': 'adsorb',  
 'adsorbing': 'adsorb',  
 'adsorbs': 'adsorb',  
 'adsorbent': 'adsorbent',  
 'adsorbents': 'adsorbent',  
 'adult': 'adult',  
 'adults': 'adult',  
 'adulterate': 'adulterate',  
 'adulterated': 'adulterate',  
 'adulterates': 'adulterate',  
 'adulterating': 'adulterate',  
 'adulterer': 'adulterer',  
 'adulterers': 'adulterer',  
 'adumbrate': 'adumbrate',  
 'adumbrated': 'adumbrate',  
 'adumbrates': 'adumbrate',  
 'adumbrating': 'adumbrate',  
 'advance': 'advance',  
 'advanced': 'advance',  
 'advances': 'advance',  
 'advancing': 'advance',  
 'advancement': 'advancement',  
 'advancements': 'advancement',  
 'advantage': 'advantage',  
 'advantaged': 'advantage',  
 'advantages': 'advantage',  
 'adventure': 'adventure',  
 'adventured': 'adventure',  
 'adventures': 'adventure',  
 'adventuring': 'adventure',  
 'adventurer': 'adventurer',  
 'adventurers': 'adventurer',  
 'adverb': 'adverb',  
 'adverbs': 'adverb',  
 'adverbial': 'adverbial',  
 'adverbials': 'adverbial',  
 'adversary': 'adversary',  
 'adversaries': 'adversary',  
 'advert': 'advert',  
 'adverted': 'advert',  
 'advertising': 'advert',  
 'adverts': 'advert',  
 'advertise': 'advertise',  
 'advertised': 'advertise',  
 'advertises': 'advertise',  
 'advertising': 'advertise',  
 'advertisement': 'advertisement',  
 'advertisements': 'advertisement',  
 'advertiser': 'advertiser',  
 'advertisers': 'advertiser',  
 'advise': 'advise'.

'advised': 'advise',  
 'advises': 'advise',  
 'advising': 'advise',  
 'adviser': 'adviser',  
 'advisers': 'adviser',  
 'advisor': 'advisor',  
 'advisors': 'advisor',  
 'advocate': 'advocate',  
 'advocated': 'advocate',  
 'advocates': 'advocate',  
 'advocating': 'advocate',  
 'adze': 'adze',  
 'adzed': 'adze',  
 'adzes': 'adze',  
 'adzing': 'adze',  
 'aeon': 'aeon',  
 'aeons': 'aeon',  
 'aerate': 'aerate',  
 'aerated': 'aerate',  
 'aerates': 'aerate',  
 'aerating': 'aerate',  
 'aerial': 'aerial',  
 'aerials': 'aerial',  
 'aerobatic': 'aerobatic',  
 'aerobatics': 'aerobatic',  
 'aerobic': 'aerobic',  
 'aerobics': 'aerobic',  
 'aerodrome': 'aerodrome',  
 'aerodromes': 'aerodrome',  
 'aerodynamic': 'aerodynamic',  
 'aerodynamics': 'aerodynamic',  
 'aerogramme': 'aerogramme',  
 'aerogrammes': 'aerogramme',  
 'aeroplane': 'aeroplane',  
 'aeroplanes': 'aeroplane',  
 'aerosol': 'aerosol',  
 'aerosolled': 'aerosol',  
 'aerosols': 'aerosol',  
 'aesthete': 'aesthete',  
 'aesthetes': 'aesthete',  
 'affair': 'affair',  
 'affairs': 'affair',  
 'affect': 'affect',  
 'affected': 'affect',  
 'affecting': 'affect',  
 'affects': 'affect',  
 'affection': 'affection',  
 'affections': 'affection',  
 'affiance': 'affiance',  
 'affianded': 'affiance',  
 'affidavit': 'affidavit',  
 'affidavits': 'affidavit',  
 'affiliate': 'affiliate',  
 'affiliated': 'affiliate',  
 'affiliates': 'affiliate',  
 'affiliating': 'affiliate',  
 'affiliation': 'affiliation',  
 'affiliations': 'affiliation',  
 'affine': 'affine',  
 'affines': 'affine',  
 'affinity': 'affinity',  
 'affinities': 'affinity',  
 'affirm': 'affirm',  
 'affirmed': 'affirm'.



'affirming': 'affirm',  
'affirms': 'affirm',  
'affirmation': 'affirmation',  
'affirmations': 'affirmation',  
'affix': 'affix',  
'affixed': 'affix',  
'affixes': 'affix',  
'affixing': 'affix',  
'afflict': 'afflict',  
'afflicted': 'afflict',  
'afflicting': 'afflict',  
'afflicts': 'afflict',  
'affliction': 'affliction',  
'afflictions': 'affliction',  
'afford': 'afford',  
'afforded': 'afford',  
'affording': 'afford',  
'affords': 'afford',  
'affront': 'affront',  
'affronted': 'affront',  
'affronting': 'affront',  
'affronts': 'affront',  
'afghan': 'afghan',  
'afghans': 'afghan',  
'afghani': 'afghani',  
'afghanis': 'afghani',  
'aficionado': 'aficionado',  
'aficionados': 'aficionado',  
'aforesay': 'aforesay',  
'aforesaid': 'aforesay',  
'afrancesado': 'afrancesado',  
'afrancesados': 'afrancesado',  
'african': 'african',  
'africans': 'african',  
'afrikaner': 'afrikaner',  
'afrikaners': 'afrikaner',  
'afro': 'afro',  
'afros': 'afro',  
'after': 'after',  
'afters': 'after',  
'afternoon': 'afternoon',  
'afternoons': 'afternoon',  
'aftershock': 'aftershock',  
'aftershocks': 'aftershock',  
'afterthought': 'afterthought',  
'afterthoughts': 'afterthought',  
'age': 'age',  
'aged': 'age',  
'ageing': 'age',  
'ages': 'age',  
'aging': 'age',  
'agency': 'agency',  
'agencies': 'agency',  
'agenda': 'agenda',  
'agendas': 'agenda',  
'agent': 'agent',  
'agents': 'agent',  
'agente': 'agente',  
'agentes': 'agente',  
'ager': 'ager',  
'agers': 'ager',  
'agglomeration': 'agglomeration',  
'agglomerations': 'agglomeration',  
'agglutinate': 'agglutinate'.

'agglutinated': 'agglutinate',  
 'agglutinating': 'agglutinate',  
 'aggravate': 'aggravate',  
 'aggravated': 'aggravate',  
 'aggravates': 'aggravate',  
 'aggravating': 'aggravate',  
 'aggravation': 'aggravation',  
 'aggravations': 'aggravation',  
 'aggregate': 'aggregate',  
 'aggregated': 'aggregate',  
 'aggregates': 'aggregate',  
 'aggregating': 'aggregate',  
 'aggregation': 'aggregation',  
 'aggregations': 'aggregation',  
 'aggression': 'aggression',  
 'aggressions': 'aggression',  
 'aggressor': 'aggressor',  
 'aggressors': 'aggressor',  
 'agister': 'agister',  
 'agisters': 'agister',  
 'agitate': 'agitate',  
 'agitated': 'agitate',  
 'agitates': 'agitate',  
 'agitating': 'agitate',  
 'agitation': 'agitation',  
 'agitations': 'agitation',  
 'agitator': 'agitator',  
 'agitators': 'agitator',  
 'agm': 'agm',  
 'agms': 'agm',  
 'agnathan': 'agnathan',  
 'agnathans': 'agnathan',  
 'agnostic': 'agnostic',  
 'agnostics': 'agnostic',  
 'agonise': 'agonise',  
 'agonised': 'agonise',  
 'agonises': 'agonise',  
 'agonising': 'agonise',  
 'agonist': 'agonist',  
 'agonists': 'agonist',  
 'agonize': 'agonize',  
 'agonized': 'agonize',  
 'agonizing': 'agonize',  
 'agony': 'agony',  
 'agonies': 'agony',  
 'agree': 'agree',  
 'agreed': 'agree',  
 'agreeing': 'agree',  
 'agrees': 'agree',  
 'agreement': 'agreement',  
 'agreements': 'agreement',  
 'agriculturalist': 'agriculturalist',  
 'agriculturalists': 'agriculturalist',  
 'agrochemical': 'agrochemical',  
 'agrochemicals': 'agrochemical',  
 'agronomist': 'agronomist',  
 'agronomists': 'agronomist',  
 'ah': 'ah',  
 'ahed': 'ah',  
 'ahhed': 'ah',  
 'ahing': 'ah',  
 'ahs': 'ah',  
 'ai': 'ai',  
 'ais': 'ai'

'aid': 'aid',  
 'aided': 'aid',  
 'aiding': 'aid',  
 'aids': 'aid',  
 'aide': 'aide',  
 'aides': 'aide',  
 'ail': 'ail',  
 'ailed': 'ail',  
 'ailing': 'ail',  
 'ails': 'ail',  
 'aileron': 'aileron',  
 'ailerons': 'aileron',  
 'ailment': 'ailment',  
 'ailments': 'ailment',  
 'aim': 'aim',  
 'aimed': 'aim',  
 'aiming': 'aim',  
 'aims': 'aim',  
 'air': 'air',  
 'aired': 'air',  
 'airing': 'air',  
 'airs': 'air',  
 'airbag': 'airbag',  
 'airbags': 'airbag',  
 'airbrake': 'airbrake',  
 'airbraked': 'airbrake',  
 'airbrakes': 'airbrake',  
 'aircondition': 'aircondition',  
 'airconditioned': 'aircondition',  
 'airconditioning': 'aircondition',  
 'aircrew': 'aircrew',  
 'aircrews': 'aircrew',  
 'airdrop': 'airdrop',  
 'airdrops': 'airdrop',  
 'airfield': 'airfield',  
 'airfields': 'airfield',  
 'airframe': 'airframe',  
 'airframes': 'airframe',  
 'airlift': 'airlift',  
 'airlifted': 'airlift',  
 'airlifting': 'airlift',  
 'airlifts': 'airlift',  
 'airline': 'airline',  
 'airlines': 'airline',  
 'airliner': 'airliner',  
 'airliners': 'airliner',  
 'airlock': 'airlock',  
 'airlocks': 'airlock',  
 'airman': 'airman',  
 'airmen': 'airman',  
 'airplane': 'airplane',  
 'airplanes': 'airplane',  
 'airport': 'airport',  
 'airports': 'airport',  
 'airship': 'airship',  
 'airships': 'airship',  
 'airshow': 'airshow',  
 'airshows': 'airshow',  
 'airstrip': 'airstrip',  
 'airstrips': 'airstrip',  
 'airwave': 'airwave',  
 'airwaves': 'airwave',  
 'airway': 'airway',  
 'airways': 'airway'.

'aisle': 'aisle',  
 'aisles': 'aisle',  
 'aitch': 'aitch',  
 'aitched': 'aitch',  
 'aitches': 'aitch',  
 'alarm': 'alarm',  
 'alarmed': 'alarm',  
 'alarming': 'alarm',  
 'alarms': 'alarm',  
 'albanian': 'albanian',  
 'albanians': 'albanian',  
 'albatross': 'albatross',  
 'albatrosses': 'albatross',  
 'albican': 'albican',  
 'albicans': 'albican',  
 'album': 'album',  
 'albums': 'album',  
 'albuman': 'albuman',  
 'albumen': 'albuman',  
 'alchemist': 'alchemist',  
 'alchemists': 'alchemist',  
 'alcohol': 'alcohol',  
 'alcohols': 'alcohol',  
 'alcoholic': 'alcoholic',  
 'alcoholics': 'alcoholic',  
 'alcove': 'alcove',  
 'alcoves': 'alcove',  
 'alder': 'alder',  
 'alders': 'alder',  
 'alderman': 'alderman',  
 'aldermen': 'alderman',  
 'ale': 'ale',  
 'aled': 'ale',  
 'ales': 'ale',  
 'alehouse': 'alehouse',  
 'alehouses': 'alehouse',  
 'alert': 'alert',  
 'alerted': 'alert',  
 'alerting': 'alert',  
 'alerts': 'alert',  
 'alga': 'alga',  
 'algae': 'alga',  
 'algerian': 'algerian',  
 'algerians': 'algerian',  
 'algorithm': 'algorithm',  
 'algorithms': 'algorithm',  
 'aliase': 'aliase',  
 'aliases': 'aliase',  
 'aliasing': 'aliase',  
 'alibi': 'alibi',  
 'alibis': 'alibi',  
 'alien': 'alien',  
 'aliens': 'alien',  
 'alienate': 'alienate',  
 'alienated': 'alienate',  
 'alienates': 'alienate',  
 'alienating': 'alienate',  
 'alight': 'alight',  
 'alighted': 'alight',  
 'alighting': 'alight',  
 'alights': 'alight',  
 'align': 'align',  
 'aligned': 'align',  
 'aligning': 'align'

```

'aligns': 'align',
'alignment': 'alignment',
'alignments': 'alignment',
'aliquot': 'aliquot',
'aliquots': 'aliquot',
'alkali': 'alkali',
'alkalis': 'alkali',
'alkaloid': 'alkaloid',
'alkaloids': 'alkaloid',
'alkane': 'alkane',
'alkanes': 'alkane',
'allay': 'allay',
'allayed': 'allay',
'allaying': 'allay',
'allays': 'allay',
'allegation': 'allegation',
'allegations': 'allegation',
'allege': 'allege',
'alleged': 'allege',
'alleges': 'allege',
'alleging': 'allege',
'allegiance': 'allegiance',
'allegiances': 'allegiance',
'allegory': 'allegory',
'allegories': 'allegory',
'allele': 'allele',
'alleles': 'allele',
'allergen': 'allergen',
'allergens': 'allergen',
'allergy': 'allergy',
'allergies': 'allergy',
'alleviate': 'alleviate',
'alleviated': 'alleviate',
'alleviates': 'alleviate',
'alleviating': 'alleviate',
'alley': 'alley',
'alleys': 'alley',
'alleyway': 'alleyway',
'alleyways': 'alleyway',
'alliance': 'alliance',
...}

```

```

In [8]: # the lemma of "skiing" is "ski"
lemmas['skiing']

```

```

Out[8]: 'ski'

```

## Question 3

Create a function that takes a list of tokens as a parameter and removes any tokens that are in `stopwords`. Test it with your `tokens` variable.

```

In [19]: # put your answer here
def clean_tokens(tokens, stopwords):

    cleaned_tokens=[]
    for term in tokens:
        if term not in stopwords:
            cleaned_tokens.append(term)
        else:
            # remove term from tokens
            tokens.remove(term)

```

```
return cleaned_tokens
```

```
cleaned_tokens=clean_tokens(tokens, stopwords)  
print(cleaned_tokens)
```

```
remove from tokens: how  
remove from tokens: are  
['Hello', 'there,', 'you?']
```

## Question 4

Create a function that takes a list of tokens as a parameter and replaces them with their lemma if that token exists in the `lemmas` variable. Test it with your `tokens` variable.

In [11]: *# put your answer here*

```
def lemmatized_tokens(tokens, lemmas):  
  
    lemmatized_tokens=[]  
    for term in tokens:  
        if term in lemmas:  
            lemmatized_term=lemmas[term]  
        else:  
            lemmatized_term = term  
        lemmatized_tokens.append(lemmatized_term)  
    return(lemmatized_tokens)  
  
lemmatized_tokens=lemmatized_tokens(tokens, lemmas)  
print(lemmatized_tokens)  
  
['Hello', 'there,', 'how', 'be', 'you?']
```

## Question 5

Discuss the pros and cons of lemmatization.

*put your answer here*

Lemmatization reduces words to their base form, providing us a more accurate results in natural language processing. Also, by converting words to base forms, it maintains data consistency and reduces dimensionality when having a huge dataset. However, in some cases, lemmatization can result in a loss of information. For example, the distinction between different verb tenses or forms may be lost, which could be important in certain analyses.