Week 2 - Variables

Author: Johanne Sejrskild
Date: 10.08.2025

Good afternoon
Today we are going to work with variables and text as data. The green excercises will be highly linked to what you did yesterday with Anna. If you find them challenging use Annas powerpoint as a help or ask. If you want to challenge yourself, try and do them all without using any help. In the yellow and red excercises we will be working with text as data. Here we will work with litterary classics and exploring the language used.

Structure of the notebook:
Green excercises

- Variables and storing data in them
- Container variable types
- Accessing elements
- Quiz

Yellow excercises

- Build a lexicon of a word in a book of you choosing
- Compare lengths of the books

Red excercises

- Find the most frequent words
- Compare pronouns used in books written by male and female authors

Start with the first excercise, and then continue in order. Feel free to work together, and see how far you can get.
The important thing is to learn, not to solve all the challenges!

Green excercises

Variables and storing data in them

Fill in the blanks:

Which variable does these data types belong to?
Whole numbers = integer Decimal numbers = float Collections of characters = string

True or False values = boolean

```python
In [ ]:  # Add some values to the variables below and test if you are correct.
         # you can use the type() function to check the data type of a variable.

         whole = 1
         decimal = 1.1
         characters = "hello"
         TrueFalse = True

         dict_var = {"Whole Numbers": whole,
                     "Decimal Numbers": decimal,
                     "Collections og characters": characters,
                     "True or False values": TrueFalse}

         for variable in dict_var:
             print(f"The variable '{variable}' is type: {type(dict_var[variable])}")
```

```
The variable 'Whole Numbers' is type: <class 'int'>
The variable 'Decimal Numbers' is type: <class 'float'>
The variable 'Collections og characters' is type: <class 'str'>
The variable 'True or False values' is type: <class 'bool'>
```

Container variable types

Python provides several built-in container types, including lists, dictionaries, tuples, and sets. These containers can store collections of data and are essential for effective data manipulation.

Lists are ordered collections of items that can be changed (mutable). Lists are defined by square brackets [].
Dictionaries store key-value pairs, are unordered, and mutable. Defined by curly braces {}.
Tuples are ordered collections of items that cannot be changed (immutable). Defined by parentheses ().
Sets are unordered collections of unique items, mutable but items must be immutable. Defined by curly braces {} or the set() function.

```python
In [ ]: # create a list, dictionary, tuple and a set.

# list
groceries = ["salat", "yoghurt"]

# dictionary
ordbog = {"kurv": "salat", "skab": "yoghurt"}

# tuple
tuplezzz = ("hej", "jeg")

# set
sæt = {"hej", "jeg"}
```

Acessing elements

Access a value from each container variable you just created

```python
In [ ]: # list — Access elements by index, starting at 0.
print(groceries[0])

# dictionary — Access values by their key.
ordbog["skuffe"] = "slik"  # Change the value associated with the key "kurv"
print(ordbog["skuffe"])

# tuple — Access elements by index, just like lists.
print(tuplezzz[0])

# set — Cannot access items by index because sets are unordered. but you can
print("hej" in sæt)
```

```
salat
slik
hej
True
```

Quiz

1. Which of the following correctly adds a new element "orange" to the fruits list?

fruits.add(orange)

fruits.append("orange")

fruits.insert{"orange"}

fruits += ["orange"]

Answer: fruits.append("orange")

2. How do you add a new key-value pair "gender": "female" to the person dictionary?

person.add("gender" = "female")

person["gender"] = "female"
person.append("gender": "female")
person.insert("gender" = "female")

Answer: person["gender"] = "female"

3. Which of the following statements about tuples is true?

Tuples are mutable.
You can add new elements to a tuple.
You access tuple elements by key.
Tuples are ordered and immutable.

Answer: Tuples are ordered and immutable

4. What will be the result of adding an existing element to a set?

The set will add another instance of the element.
The set will remain unchanged because it only holds unique elements.
The operation will result in an error.
The set will reorder its elements.

Answer: The operation will result in an error

Yellow excercises

Before you start on the *yellow excercises* you have to install and import the package containing all the litterary classics.
You will also get an example of how to build a lexicon of the use of the word "whale" in Moby Dick and the words surrounding it.

So:
Run each cell underneath one at a time, in order. If something in one cell doesn't work right, it might be because you have overwritten a variable, so try going back and running all the previous cells again.

In [ ]:
```
# install the natural language toolkit package (nltk), which has a copy of t

%pip install nltk
```

```
Defaulting to user installation because normal site-packages is not writeabl
e
Collecting nltk
  Downloading nltk-3.9.1-py3-none-any.whl (1.5 MB)
       |████████████████████████████████| 1.5 MB 3.5 MB/s eta 0:00:01
Collecting regex>=2021.8.3
  Downloading regex-2025.9.1-cp39-cp39-macosx_11_0_arm64.whl (286 kB)
       |████████████████████████████████| 286 kB 9.8 MB/s eta 0:00:01
Collecting click
  Downloading click-8.1.8-py3-none-any.whl (98 kB)
       |████████████████████████████████| 98 kB 9.0 MB/s  eta 0:00:01
Collecting tqdm
  Downloading tqdm-4.67.1-py3-none-any.whl (78 kB)
       |████████████████████████████████| 78 kB 18.5 MB/s eta 0:00:01
Collecting joblib
  Downloading joblib-1.5.2-py3-none-any.whl (308 kB)
       |████████████████████████████████| 308 kB 30.0 MB/s eta 0:00:01
Installing collected packages: tqdm, regex, joblib, click, nltk
Successfully installed click-8.1.8 joblib-1.5.2 nltk-3.9.1 regex-2025.9.1 tq
dm-4.67.1
WARNING: You are using pip version 21.2.4; however, version 25.2 is availabl
e.
You should consider upgrading via the '/Library/Developer/CommandLineTools/u
sr/bin/python3 -m pip install --upgrade pip' command.
Note: you may need to restart the kernel to use updated packages.
```

In [ ]:
```python
import nltk
```

In [ ]:
```python
# import the nltk package so that it is accessible to Python, and download a
nltk.download('gutenberg')
```

```
[nltk_data] Downloading package gutenberg to /Users/sofiascharf-
[nltk_data]     matthiesen/nltk_data...
[nltk_data]   Unzipping corpora/gutenberg.zip.
```

Out[ ]: True

Now we have installed and imported the books, but we want to extract one and do a bit
of data cleeaning so we can look at the words used.

I have chosen to look at Moby Dick so lets find that

In [ ]:
```python
# Create a variable called "macbeth" which contains the text of the book.
mobydick = nltk.corpus.gutenberg.raw('melville-moby_dick.txt')

# make all characters lowercase
mobydick = mobydick.lower()

# remove the "\n" and "\r" characters, which indicate line breaks in the tex
mobydick = mobydick.replace('\n', ' ')
mobydick = mobydick.replace('\r', ' ')

# split up the text into a long list of individual words
mobydick = mobydick.split()
```

Question: Why is it necessary to split the text into a list of words in order to do analysis

on them?

Otherwise my lil computer won't know that they are not the same item

In [ ]:
```python
# make a variable called "concordance", and fill it with every occurrence of
lexicon = []

for i, val in enumerate(mobydick):
    if val == "whale":
            lexicon.append(str(' '.join(mobydick[i-5:i+5])))
```

In [ ]:
```python
# take a look at what the algorithm has found
lexicon
```

```
Out[ 1: ['at least, take the higgledy-piggledy whale statements, however authentic,
         in',
          'that monstrous bulk of the whale or ork we have',
          "paine, like as the wounded whale to shore flies thro'",
          '--dryden\'s annus mirabilis. "while the whale is floating at the',
          "jonas-in-the-whale. ... some say the whale can't open his mouth,",
          'i was told of a whale taken near shetland, that',
          'that he caught once a whale in spitzbergen that was',
          'and the breath of the whale is frequendy attended with',
          'contemptible in the comparison. the whale is doubtless the largest',
          'iceland in 1772. "the spermacetti whale found by the nantuckois,',
          'in the fishermen." --thomas jefferson\'s whale memorial to the french',
          'the nantucket whale-fishery. "spain--a great whale stranded on the shore
         s',
          'to royal fish, which are whale and sturgeon. and these,',
          'on high, up-spouted by a whale in air, to express',
          'one.) "the aorta of a whale is larger in the',
          'whale\'s heart." --paley\'s theology. "the whale is a mammiferous anima
         l',
          'purpose of extending the spermaceti whale fishery. "in the free',
          "people's king. not a mightier whale than this in the",
          'had been killed by a whale in the pacific ocean,',
          'of the shipwreck of the whale ship essex of nantucket,',
          'destroyed by a large sperm whale in the pacific ocean."',
          'english miles. ... "sometimes the whale shakes its tremendous tail',
          'fresh attacks, the infuriated sperm whale rolls over and over;',
          'the known species of the whale tribe." --frederick debell bennett\'s',
          'ahoy! do you see that whale now?" "ay ay, sir!',
          'being once pursued by a whale which he had wounded,',
          'breakwater at nantucket. 1828. "the whale fell directly over him,',
          'him in a moment." --"the whale and his captors, or',
          'of whales." --tales of a whale voyager to the arctic',
          'they departed." --cruise in a whale boat. "suddenly a mighty',
          'while." --miriam coffin or the whale fisherman. "the whale is',
          'or the whale fisherman. "the whale is harpooned to be',
          'jaws of a large sperm whale close to the head',
          'for your lives!\'" --wharton the whale killer. "so be cheery,',
          'overwhelming idea of the great whale himself. such a portentous',
          'where the first dead american whale was stranded. where else',
          'was killed by a sperm whale on the coast of',
          'my complaints-- no more the whale did me confine. "with',
          'that stifling hour, when the whale shall hold him in',
          'jaws awaiting him; and the whale shoots-to all his ivory',
          "the belly of hell'--when the whale grounded upon the ocean's",
          'blackness of the sea, the whale came breeching up towards',
          'side by side the world-wandering whale ships lay silent and',
          'him? well, spose him one whale eye, well, den!" and',
          'him whale-e eye; why, dad whale dead." "quick, bildad," said',
          "in the green locker! don't whale it too much a'",
          'append the initials s.w.f. (sperm whale fishery) to his visiting',
          'the world scouts at us whale hunters, yet does it',
          'oh, the world! until the whale fishery rounded cape horn,',
          'split helmet every time. the whale has no famous author,',
          'chronicler, you will say. the whale no famous author, and',
          'old english statutory law, the whale is declared "a royal',
          "oh, that's only nominal! the whale himself has never figured",
```

```
'any grand imposing way. the whale never figured in any',
'stubb, or almost any other whale hunter. starbuck was no',
'his poor opinion, the wondrous whale was but a species',
'mast employed in the american whale fishery, are americans born,',
'the same with the american whale fishery as with the',
'some systematized exhibition of the whale in his broad genera,',
'names in this list of whale authors, only those following',
'compared with which the greenland whale is almost unworthy mentioning.',
'it said, that the greenland whale is an usurper upon',
'ye! good people all,--the greenland whale is deposed,--the great sperm',
'whale is deposed,--the great sperm whale now reigneth! there are',
'to put the living sperm whale before you, and at',
'original matter touching the sperm whale to be found in',
'a moot point whether a whale be a fish. in',
'old fashioned ground that the whale is a fish, and',
'what internal respect does the whale differ from other fish.',
'to be short, then, a whale is a spouting fish',
'above definition of what a whale is, i do by',
'creature hitherto identified with the whale by the best informed',
'grand divisions of the entire whale host. *i am aware',
'centuries ago, when the sperm whale was almost wholly unknown',
'quickening humor of the greenland whale which the first syllable',
'to be bestowed upon the whale from which this spermaceti',
'the english naturalists; the greenland whale of the english whalemen;',
'the swedes. it is the whale which for more than',
'arctic seas; it is the whale which the american fishermen',
'world, designated by them right whale cruising grounds. some pretend',
'a difference between the greenland whale of the english and',
'the english and the right whale of the americans. but',
'so repellingly intricate. the right whale will be elsewhere treated',
'sea and is commonly the whale whose distant jet is',
'essential particulars. thus, the sperm whale and the humpbacked whale,',
'ceases. then, this same humpbacked whale and the greenland whale,',
'i. (folio) chapter iv. (hump-back).--this whale is often seen on',
'distinguish him, since the sperm whale also has a hump',
'(folio), chapter v. (razor-back).--of this whale little is known but',
'grin on his face. this whale averages some sixteen or',
'more profitably employed, the sperm whale hunters sometimes capture the',
'(octavo), chapter iv. (killer).--of this whale little is precisely know
n',
'my definition of what a whale is--i.e. a spouting fish,',
'then see the great sperm whale himself in miniature. book',
'ago, the command of a whale ship was not wholly',
'no exception to most american whale captains, who, as a',
'the hope of capturing one whale more. now, as the',
'for a sail, or a whale just bearing in sight.',
'the early times of the whale fishery, ere ships were',
'the mast-heads of a southern whale ship are unprovided with',
'thou hast not raised a whale yet. whales are scarce',
'pull to, men?" "a dead whale or a stove boat!"',
'ye raises me a white-headed whale with a wrinkled brow',
'ahab," said tashtego, "that white whale must be the same',
'"do ye know the white whale then, tash?" "does he',
'it was that accursed white whale that razeed me; made',
'men! to chase that white whale on both sides of',
'wall? to me, the white whale is that wall, shoved',
```

```
                    'hate; and be the white whale agent, or be the',
                    'agent, or be the white whale principal, i will wreak',
                    'tide flow wide. the hated whale has the round watery',
                    'the sharkish sea. the white whale is their demigorgon. hark!',
                    'only, the unaccompanied, secluded white whale had haunted those unciviliz
                 ed',
                    'mostly frequented by the sperm whale fishermen. but not all',
                    'such a meridian, a sperm whale of uncommon magnitude and',
                    'presumption, i say, that the whale in question must have',
                    'as of late the sperm whale fishery had been marked',
                    'the perils of the sperm whale fishery at large, than',
                    'encounter between ahab and the whale had hitherto been popularly',
                    'him, as for any other whale of that species. but',
                    'the story of the white whale had eventually come. nor',
                    'in this matter, so the whale fishery surpasses every other',
                    'outblown rumors of the white whale did in the end',
                    'tremendousness of the great sperm whale anywhere more feelingly comprehen
                 ded,',
                    'naturalists--olassen and povelson--declaring the sperm whale not only to
                 be',
                    'earlier days of the sperm whale fishery, when it was',
                    'an apparition as the sperm whale was not for mortal',
                    'be linked with the white whale in the minds of',
                    'hidden ways of the sperm whale when beneath the surface',
                    'repeated, intrepid assaults, the white whale had escaped alive; it',
                    'that so much invested the whale with natural terror, as',
                    'and spiritual exasperations. the white whale swam before him as',
                   "chasing with curses a job's whale round the world, at",
                    'seemed almost theirs; the white whale as much their insufferable',
                    'came to be--what the white whale was to them, or',
                    'the whale. what the white whale was to ahab, has',
                    'was the whiteness of the whale that above all things',
                    'all these things the albino whale was the symbol. wonder',
                    'one voyage of the entire whale fleet carefully collated, then',
                    'the migrations of the sperm whale would be found to',
                    'direction taken by any one whale be straight as a',
                    'deadly encounters with the white whale had taken place; there',
                    'of unknown regions. meanwhile, the whale he had struck must',
                    'purpose. this man and this whale again came together, and',
                    'well known in the sperm whale fishery, however ignorant the',
                    'historical instances where a particular whale in the ocean has',
                    'popularly cognisable. why such a whale became thus marked was',
                    'in that respect any chance whale may be, they soon',
                    'of perilousness about such a whale as there did about',
                    'some indefinite idea that a whale is an enormous creature',
                    'point is this: the sperm whale is in some cases',
                    'what is more, the sperm whale has done it. first:',
                    'when, suddenly, a very large whale escaping from the boats,',
                    'on the part of the whale (many of which impressions',
                    'to encounter, though from the whale hunters i have now',
                    'denied for example, that any whale could so smite his',
                   "the commodore's interview with that whale as providential. was not",
                    'i tell you, the sperm whale will stand no nonsense.',
                    'all caused by an unseen whale vertically bumping the hull',
                    'attached to a running sperm whale have, in a calm,',
                    'ship, and secured there; the whale towing her great hull',
```

'i fancied that the sperm whale had been always unknown',
'the presence of the sperm whale in the mediterranean. i',
'the dardanelles, hence a sperm whale could, by the same',
'his vindictiveness towards the white whale might have possibly extended',
'chances that each subsequently encountered whale would prove to be',
'would elapse ere the white whale was seen. during that',
'sordidness. granting that the white whale fully incites the hearts',
'all was commotion. the sperm whale blows as a clock',
'times evinced by the sperm whale when, sounding with his',
'men! it ain\'t the white whale to-day! give way!" now',
'boat lowerings ere the white whale had torn him. all',
'for catching turns with the whale line. its top is',
'pequod, and the great white whale its object. "queequeg," said',
'going plump on a flying whale with your sail set',
'backing water up to a whale face foremost. ha, ha!',
'face foremost. ha, ha! the whale would give them squint',
'of going on to the whale i must resign my',
"starbuck's driving on to his whale almost in the teeth",
'of detail; so has the whale fishery. in a pirate,',
'general interest in the white whale was now wildly heightened',
'obscurely to involve with the whale a certain wondrous, inverted',
'whale-cry upon first sighting a whale from the mast-head, still',
'so suddenly perceiving the snowy whale within fifty yards of',
'of moby dick. but the whale rushed round in a',
'he cut it; and the whale was free. but, at',
'gave chase again; but the whale eluded them, and finally',
'dreams sees the awful white whale that destroyed him. "\'are',
'the true form of the whale as he actually appears',
'his own absolute body the whale is moored alongside the',
'proving such pictures of the whale all wrong. it may',
'there shadowed forth. the hindoo whale referred to, occurs in',
"these? as for the book-binder's whale winding like a vine-stalk",
'is made of representing the whale with perpendicular flukes. then',
'purpose of extending the spermaceti whale fisheries." in this book',
'make the eye of that whale a bow-window some five',
'seem inelegant, but this unsightly whale looks much like an',
'naturalist, published a scientific systemized whale book, wherein are sev
eral',
'of the mysticetus or greenland whale (that is to say,',
"a word, frederick cuvier's sperm whale is not a sperm",
'manifold mistakes in depicting the whale are not so very',
'contour between a young sucking whale and a full-grown platonian',
'the mere skeleton of the whale bears the same relation',
'artificial covering. "however recklessly the whale may sometimes serve u
s,"',
'finding out precisely what the whale really looks like. and',
"all beale's drawings of this whale are good, excepting the",
'effect. some of the sperm whale drawings in j. ross',
'truthful idea of the living whale as seen by his',
'first engraving a noble sperm whale is depicted in full',
'crew are scattered about the whale in contrasting expressions of',
'and maccaroni, which the right whale sometimes carries on his',
'the real spirit of the whale hunt. for the most',
'part, the english and american whale draughtsmen seem entirely content',
'leviathanic life, with a right whale alongside; the vessel (in',
'the torments of the boiling whale is going up like',

```
                    'porter is sleepy, the anvil-headed whale would be best. but',
                    'would i could mount that whale and leap the topmost',
                    'substance, upon which the right whale largely feeds. for leagues',
                    'those latitudes, where the right whale is often chased. but',
                    'first sight of the particular whale he pursued; however this',
                    'to furnish to the sperm whale his only food. for',
                    'act of feeding, the spermaceti whale obtains his whole food',
                    'boat, in case the stricken whale should sound so deep',
                    'harpoon. in these instances, the whale of course is shifted',
                    'the boat, and were the whale then to run the',
                    'fathoms off, a gigantic sperm whale lay rolling in the',
                    'spouting his vapoury jet, the whale looked like a portly',
                    'his sounding had elapsed, the whale rose again, and being',
                    'was obvious, now, that the whale had at length become',
                    'way, till at length the whale somewhat slackened his flight.',
                    'the bowsman, as the waning whale relaxed in his wrath.',
                    'some gold watch that the whale might have swallowed, and',
                    'abating in his flurry, the whale once more rolled out',
                    'instant, that is, when the whale starts to run, the',
                    'much the speed of the whale as the before described',
                    'violent, convulsive running of the whale upon receiving the first',
                    'secure it again until the whale is fairly captured and',
                    "chapter 64 stubb's supper. stubb's whale had been killed some",
                    'bulwarks. vacantly eyeing the heaving whale for a moment, he',
                    'tail to the bows, the whale now lies with its',
                    'the ship has upon the whale when moored alongside, is',
                    'somewhat intemperately fond of the whale as a flavorish thing',
                    'particular part of the sperm whale designated by stubb; comprising',
                    'huge globular pieces of the whale of the bigness of',
                    'right to dat whale; dat whale belong to some one',
                    'breakfast--don\'t forget." "wish, by gor! whale eat him, \'stead of',
                    'his hammock. chapter 65 the whale as a dish. that',
                    'the tongue of the right whale was esteemed a great',
                    'his hunters at least, the whale would by all hands',
                    'but what further depreciates the whale as a civilized dish,',
                    'case of a small sperm whale the brains are accounted',
                    'not, perhaps, entirely because the whale is so excessively unctuous',
                    'but stubb, he eats the whale by its own light,',
                    'as the blubber envelopes the whale precisely as the rind',
                    'the windlass continually keeps the whale rolling over and over',
                    'hoisting and lowering simultaneously; both whale and windlass heaving, th
e',
                    'proper skin of the tremendous whale is thinner and more',
                    'visible surface of the sperm whale is not the least',
                    'the hieroglyphics upon one sperm whale in particular, i was',
                    'mystic rocks, too, the mystic-marked whale remains undecipherable. this a
llusion',
                    'the exterior of the sperm whale presents, he not seldom',
                    'a little resemble the sperm whale in this particular. it',
                    'that such scratches in the whale are probably made by',
                    'happy and significant. for the whale is indeed wrapt up',
                    'of his body, that the whale is enabled to keep',
                    'fire; whereas, like man, the whale has lungs and warm',
                    'the blood of a polar whale is warmer than that',
                    'white body of the beheaded whale flashes like a marble',
                    'from which not the mightiest whale is free. nor is',
```

```
'the beheading of the sperm whale is a scientific anatomical',
'anatomical feat, upon which experienced whale surgeons very much pride',
'without reason. consider that the whale has nothing that can',
'it belong to a small whale it is hoisted on',
"in jewellers' scales. the pequod's whale being decapitated and the",
'the ships of the american whale fleet have each a',
'provided with it. thereby, the whale commanders are enabled to',
'gibbering insanity, pronouncing the white whale to be no less',
'harpooneer shall remain on the whale till the whole tensing',
'flounders about, half on the whale and half in the',
'and then from between the whale and ship, which would',
'poor queequeg from between the whale and the ship--where he',
'except when i hit a whale or something of that',
'yet now that a sperm whale had been brought alongside',
'was made that a right whale should be captured that',
'in the tubs, and the whale not sounding very rapidly,',
'on the water, while the whale beyond also rose to',
'to fly. but the fagged whale abated his speed, and',
"with him, i'll eat this whale in one mouthful. but",
'ahead, and let\'s get the whale alongside." "i think i',
'here hailed, to tow the whale on the larboard side,',
'of folio leviathans, the sperm whale and the right whale',
'sperm whale and the right whale are by far the',
'but the ear of the whale is full as curious',
'be observed between the sperm whale and the right. while',
'vast a being as the whale should see the world',
"like a ship's jib-boom. this whale is not dead; he",
'marvellous manner. but if this whale be a king, he',
'pity, now, that this unfortunate whale should be hare-lipped. the',
'fibres, through which the right whale strains the water, and',
'greater age to the right whale than at first glance',
'reminds us that the right whale really has a sort',
'i started with--that the sperm whale and the right whale',
'sperm whale and the right whale have almost entirely different',
"whale's. nor in the sperm whale are there any of",
'a tongue. again, the right whale has two external spout-holes,',
'two external spout-holes, the sperm whale only one. look your',
'in facing death? this right whale i take to have',
'moreover you observe that the whale has no external nose;',
"the forehead of the sperm whale were paved with horses'",
'this; that though the sperm whale stove a passage through',
'so the tun of the whale contains by far the',
'heidelburgh tun of the sperm whale embraces the entire length',
'can. physiognomically regarded, the sperm whale is an anomalous creatur
e.',
'grandeur. a nose to the whale would have been impertinent.',
'sperm whale? has the sperm whale ever written a book,',
'that had the great sperm whale been known to the',
'is tongueless; and the sperm whale has no tongue, or',
'high seat, the great sperm whale shall lord it. champollion',
'the nut. if the sperm whale be physiognomically a sphinx,',
'peremptorily deny that the sperm whale has any other brain',
'some other infirmity. whether this whale belonged to the pod',
'astern; but look, did ever whale yaw so before? it',
'way; so did this old whale heave his aged bulk,',
'pitiable, and maddening sight. the whale was now going head',
```

'endless inclined plane! hurrah! this whale carries the everlasting mai
l!"',
 'in the air. and the whale soon ceasing to sound,',
 'that the longer the stricken whale stays under water, the',
 'him––in a full grown sperm whale something less than 2000',
 'how appalling to the wounded whale must have been such',
 'the boats, and soon the whale broke water within two',
 'more than sufferable anguish, the whale now spouting thick blood,',
 'was a buoy; the sunken whale being suspended a few',
 'the ship drew nigh, the whale was transferred to her',
 'and so low had the whale now settled that the',
 'of the recently killed sperm whale is a very curious',
 'it. usually the dead sperm whale floats with great buoyancy,',
 'said, however, that the sperm whale is far less liable',
 'from this incumbrance the sperm whale is wholly free. but',
 'or several days, the sunken whale again rises, more buoyant',
 'new zealand, when a right whale gives token of sinking,',
 'it said, that the first whale attacked by our brotherhood',
 'the true form of the whale was unknown to artists;',
 "in perseus' case, st. george's whale might have crawled up",
 'had to do with a whale like their great patron),',
 'whaleman; at any rate the whale caught him, if he',
 'story of hercules and the whale is considered to be',
 'story of jonah and the whale in the preceding chapter.',
 "one of which represented jonah's whale with two spouts in",
 'on second thoughts, the right whale is toothless. another reason',
 'by, some vessel with a whale for a figure–head; and,',
 'who have opined that the whale mentioned in the book',
 'jonah was swallowed by the whale in the mediterranean sea,',
 'no other way for the whale to land the prophet',
 'being too shallow for any whale to swim in. besides,',
 'in fleecy foam, the towing whale is forty feet ahead.',
 'in skilful leash. the agonized whale goes into his flurry;',
 "like a human being's, the whale can only live by",
 'to the surface, the sperm whale will continue there for',
 'alike. now, why should the whale thus insist upon having',
 'he will. but the sperm whale only breathes about one',
 'has been said that the whale only breathes through his',
 'is, nevertheless, that the sperm whale has no proper olfactories.',
 'exclusion of water, therefore the whale has no voice; unless',
 'then again, what has the whale to say? seldom have',
 'the spout of the sperm whale is the mere vapour',
 'all. and as for this whale spout, you might almost',
 'are close enough to a whale to get a close',
 'the desert; even then, the whale always carries a small',
 'the precise nature of the whale spout. it will not',
 'measureless force of the whole whale seems concentrated to a',
 'significant, that while one sperm whale only fights another sperm',
 'whale only fights another sperm whale with his head and',
 'the light buoyancy of the whale boat, and the elasticity',
 'to me, that in the whale the sense of touch',
 'when in maidenly gentleness the whale with a certain soft',
 'pity it is that the whale does not possess this',
 'fourth: stealing unawares upon the whale in the fancied security',
 'in this chapter, between the whale and the elephant, so',
 'of general bulk between the whale and the elephant is',

```
                        'the same respect to the whale that a dog does',
                        'wanting other motions of the whale in his general body,',
                        'almost all the known sperm whale cruising grounds of the',
                        'this aggregation of the sperm whale into such immense caravans,',
                        'forward-slanting spout of the sperm whale presents a thick curled',
                        'making for some one lone whale on the outskirts of',
                        'on the part of the whale struck under such circumstances,',
                        'as, blind and deaf, the whale plunged forward, as if',
                        'drew out, and the towing whale sideways vanished; then, with',
                        'moisture thrown off by the whale in his more quiet',
                        'the final spring, the unborn whale lies bent like a',
                        'precious parts in a nursing whale are cut by the',
                        'custom when fast to a whale more than commonly powerful',
                        'hauling it back again. a whale wounded (as we afterwards',
                        'accidents of the fishery, this whale had become entangled in',
                        'shove him off, you queequeg--the whale there!--prick him!--hit him! stan
                      d',
                       'sights are seen, my lord whale keeps a wary eye',
                       'might be named, my lord whale has no taste for',
                       'isolation to which the schoolmaster whale betakes himself in his',
                       'laws and regulations of the whale fishery, of which the',
                       'are cruising in company, a whale may be struck by',
                       'to claim possession of a whale previously chased or killed',
                       'a hard chase of a whale in the northern seas;',
                       'had remained attached to the whale at the time of',
                       'that the examples of the whale and the lady were',
                       'killing and beaching a fine whale which they had originally',
                       'meanwhile ruefully glancing from the whale to the stranger. but',
                       'his." in a word, the whale was seized and sold,',
                       'of the duke to the whale was a delegated one',
                       'for it. says plowdon, the whale so caught belongs to',
                       'of the greenland or right whale was largely used in',
                       'by the english law writers--the whale and the sturgeon; both',
                       'betokened that some sort of whale must be alongside. as',
                       'it was plain that the whale alongside must be what',
                       'blasted whale, that is, a whale that has died unmolested',
                       'the frenchman had a second whale alongside; and this second',
                       'whale alongside; and this second whale seemed even more of',
                       'his nose at such a whale as this, however much',
                       'breakers, mistaking them for sperm whale spouts; yes, and sometimes',
                       'with our leavings, the drugged whale there, i mean; aye,',
                      "he'll get from that drugged whale there, wouldn't be fit",
                       'fever caught from a blasted whale they had brought alongside."',
                       'by pulling out the lighter whale of the two from',
                       'benevolently towed away at his whale the other way, ostentatiously',
                       'and unloading one of these whale cemeteries, in the greenland',
                       'the blubber of the dutch whale fleet to be tried',
                       'nose. nor indeed can the whale possibly be otherwise than',
                       'shall i liken the sperm whale to for fragrance, considering',
                       'her own. now, in the whale ship, it is not',
                       'that part of the slack whale line coming against his',
                       'water. that instant the stricken whale started on a fierce',
                       'roared stubb; and so the whale was lost and pip',
                       'the likes of you; a whale would sell for thirty',
                       'line; and hence, when the whale started to run, pip',
                       'turned upon him; and the whale was winged. in three',
```

```
                    "squeeze of the hand. that whale of stubb's, so dearly",
                    'business of preparing the sperm whale for the try-works. first',
                    'self-consuming misanthrope, once ignited, the whale supplies his own fue
                 l',
                    'from the back of the whale remains clinging to the',
                    'and listen!" "if the white whale be raised, it must',
                    'a white arm of sperm whale bone, terminating in a',
                    'it, they had had a whale alongside a day or',
                    'was ignorant of the white whale at that time. well,',
                    'and what a noble great whale it was--the noblest and',
                    "i didn't then know what whale it was that had",
                    'the digestive organs of the whale are so inscrutably constructed',
                    'is yours; only let the whale have another chance at',
                    'and thus the vast sperm whale grounds of the pacific',
                    'house fitted out a discovery whale ship of their own,',
                    'at all sparing of historical whale research, when it has',
                    'english were preceded in the whale fishery by the hollanders,',
                    'in the fishery, as every whale ship must carry its',
                    'very interesting account of its whale fishery. and in this',
                    'that ancient greenland and spitzbergen whale fishery. in the first',
                    'can you land a full-grown whale on your deck for',
                    'to, a small cub sperm whale was once bodily hoisted',
                    'of a greenland or river whale in the united states."',
                    "of those parts. sir clifford's whale has been articulated throughout;",
                    'for the largest sized greenland whale of sixty feet in',
                    'calculation, i say, a sperm whale of the largest magnitude,',
                    'its fullest circumference, such a whale will weigh at least',
                    'that the skeleton of the whale is by no means',
                    'invested body of this particular whale must have been at',
                    'sea, can the fully invested whale be truly and livingly',
                    'from his mighty bulk the whale affords a most congenial',
                    'to be used by a whale author like me. one',
                    'that the skeleton of the whale furnishes but little clue',
                    'god upon the temple, no whale can pass it without',
                    'was cast forth by the whale at the base of',
                    'this afric temple of the whale i leave you, reader,',
                    '3), sets down the right whale at one hundred metres,',
                    'believe these stories? no. the whale of to-day is as',
                    'that of all animals the whale alone should have degenerated.',
                    'to show that the hunted whale cannot now escape speedy',
                    'much more may the great whale outlast all hunting, since',
                    'these things, we account the whale immortal in his species,',
                    'its rats, then the eternal whale will still survive, and',
                    'bone, and cross-beams of sperm whale ivory, the carpenter makes',
                    'in which the hated white whale must even then be',
                    'hull, "stern all! the white whale spouts thick blood!" chapter',
                    'that will stick in a whale like his own fin-bone.',
                    'lower jaw of the last whale they had slain. signals,',
                    'negroes, with glittering fiddle-bows of whale ivory, were presiding ove
                 r',
                    'sea and sky, sun and whale both stilly died together;',
                    'calm. nor has this thy whale sunwards turned his dying',
                    'my foster-brothers!" chapter 117 the whale watch. the four whales',
                    'and ere noon the dead whale was brought to the',
                    'oaths to hunt the white whale are as binding as',
                    'was concluded that the stricken whale must have indefinitely run',
```

```
              'indifference with which the white whale tore his hunters, whether',
              'the first sight of the whale myself,"--he said. "aye! ahab',
              'the fin, where the white whale most feels his accursed',
              'isle. he declared that a whale must be near. soon',
              'rigging to behold the famous whale they had so long',
              'heel. from this height the whale was now seen some',
              'could have raised the white whale first. there she blows!--there',
              'did surpass the glorified white whale as he so divinely',
              'away--on each bright side, the whale shed off enticings. no',
              'thrilled for an instant, the whale obliquely lying on his',
              'in this attitude the white whale now shook the slight',
              'in and out, as the whale dallied with the doomed',
              'described. by this motion the whale must best and most',
              'she effectually parted the white whale from his victim. as',
              'abide here till the white whale is dead; and then,',
              'the simple observation of a whale when last descried, they',
              'so many hours hence this whale will have gone two',
              'his head, did the white whale now reveal his vicinity;',
              'the furthest depths, the sperm whale thus booms his entire',
              'his immeasurable bravadoes the white whale tossed himself salmon-like t
            o',
              'them he would take the whale head-and-head,--that is, pull straight',
              'to his eye; the white whale churning himself into furious',
              'his untraceable evolutions, the white whale so crossed and recrossed,',
              'now for a moment the whale drew aside a little,',
              'again. that instant, the white whale made a sudden rush',
              'from the sea, the white whale dashed his broad forehead',
              'though!" when dusk descended, the whale was still in sight',
              'him?" cried ahab; but the whale was not yet in',
              'the palms. leeward! the white whale goes that way; look',
              'nay, to-night, when the white whale lies down there, tied',
              'the pequod since the white whale had been first descried;',
              'arm, ahab knew that the whale had sounded; but intending',
              'strained planks; and as the whale swimming out from them,',
              'during the past night, the whale had reeled the involutions',
              'swim to feast on the whale or on ahab?--but pull',
              'oblivious of its advance--as the whale sometimes will--and ahab was',
              'ungraduated, instantaneous swiftness, the white whale darted through the
            weltering',
              'of the sea-crashing boat, the whale wheeled round to present',
              'beneath the settling ship, the whale ran quivering along its',
              'harpoon was darted; the stricken whale flew forward; with igniting']
```

In [ ]:
```python
# let's see how many instances of the word "whale" were found
len(lexicon)
```

Out[ ]: 528

Let's try again, but this time let's just search for "the whale", not "whale" by itself. Try to comment each line of the code and explain what it does

In [ ]:
```python
lexicon_twowords = [] # creates new empty list

for i, val in enumerate(mobydick): # initiates for loop in enumerated list
    if val == "whale": # if the word is "whale"
```

```python
            if mobydick[i-1] == "the": # and the word before is "the"
                lexicon_twowords.append(str(' '.join(mobydick[i-5:i+5]))) # then
```

In [ ]:
```python
# take a look at what the algorithm has found
lexicon_twowords
```

```
Out[ 1]:  ['that monstrous bulk of the whale or ork we have',
           '--dryden\'s annus mirabilis. "while the whale is floating at the',
           "jonas-in-the-whale. ... some say the whale can't open his mouth,",
           'and the breath of the whale is frequendy attended with',
           'contemptible in the comparison. the whale is doubtless the largest',
           'of the shipwreck of the whale ship essex of nantucket,',
           'english miles. ... "sometimes the whale shakes its tremendous tail',
           'the known species of the whale tribe." --frederick debell bennett\'s',
           'while." --miriam coffin or the whale fisherman. "the whale is',
           'for your lives!\'" --wharton the whale killer. "so be cheery,',
           'my complaints-- no more the whale did me confine. "with',
           'that stifling hour, when the whale shall hold him in',
           'jaws awaiting him; and the whale shoots-to all his ivory',
           "the belly of hell'--when the whale grounded upon the ocean's",
           'blackness of the sea, the whale came breeching up towards',
           'oh, the world! until the whale fishery rounded cape horn,',
           'split helmet every time. the whale has no famous author,',
           'chronicler, you will say. the whale no famous author, and',
           'old english statutory law, the whale is declared "a royal',
           "oh, that's only nominal! the whale himself has never figured",
           'any grand imposing way. the whale never figured in any',
           'some systematized exhibition of the whale in his broad genera,',
           'old fashioned ground that the whale is a fish, and',
           'what internal respect does the whale differ from other fish.',
           'creature hitherto identified with the whale by the best informed',
           'to be bestowed upon the whale from which this spermaceti',
           'the swedes. it is the whale which for more than',
           'arctic seas; it is the whale which the american fishermen',
           'sea and is commonly the whale whose distant jet is',
           'the early times of the whale fishery, ere ships were',
           'presumption, i say, that the whale in question must have',
           'encounter between ahab and the whale had hitherto been popularly',
           'in this matter, so the whale fishery surpasses every other',
           'that so much invested the whale with natural terror, as',
           'was the whiteness of the whale that above all things',
           'of unknown regions. meanwhile, the whale he had struck must',
           'on the part of the whale (many of which impressions',
           'to encounter, though from the whale hunters i have now',
           'ship, and secured there; the whale towing her great hull',
           'for catching turns with the whale line. its top is',
           'face foremost. ha, ha! the whale would give them squint',
           'of going on to the whale i must resign my',
           'of detail; so has the whale fishery. in a pirate,',
           'obscurely to involve with the whale a certain wondrous, inverted',
           'of moby dick. but the whale rushed round in a',
           'he cut it; and the whale was free. but, at',
           'gave chase again; but the whale eluded them, and finally',
           'the true form of the whale as he actually appears',
           'his own absolute body the whale is moored alongside the',
           'proving such pictures of the whale all wrong. it may',
           'is made of representing the whale with perpendicular flukes. then',
           'manifold mistakes in depicting the whale are not so very',
           'the mere skeleton of the whale bears the same relation',
           'artificial covering. "however recklessly the whale may sometimes serve u
          s,"',
           'finding out precisely what the whale really looks like. and',
```

```
'crew are scattered about the whale in contrasting expressions of',
'the real spirit of the whale hunt. for the most',
'harpoon. in these instances, the whale of course is shifted',
'the boat, and were the whale then to run the',
'spouting his vapoury jet, the whale looked like a portly',
'his sounding had elapsed, the whale rose again, and being',
'was obvious, now, that the whale had at length become',
'way, till at length the whale somewhat slackened his flight.',
'some gold watch that the whale might have swallowed, and',
'abating in his flurry, the whale once more rolled out',
'instant, that is, when the whale starts to run, the',
'much the speed of the whale as the before described',
'violent, convulsive running of the whale upon receiving the first',
'secure it again until the whale is fairly captured and',
'tail to the bows, the whale now lies with its',
'the ship has upon the whale when moored alongside, is',
'somewhat intemperately fond of the whale as a flavorish thing',
'huge globular pieces of the whale of the bigness of',
'his hammock. chapter 65 the whale as a dish. that',
'his hunters at least, the whale would by all hands',
'but what further depreciates the whale as a civilized dish,',
'not, perhaps, entirely because the whale is so excessively unctuous',
'but stubb, he eats the whale by its own light,',
'as the blubber envelopes the whale precisely as the rind',
'the windlass continually keeps the whale rolling over and over',
'that such scratches in the whale are probably made by',
'happy and significant. for the whale is indeed wrapt up',
'of his body, that the whale is enabled to keep',
'fire; whereas, like man, the whale has lungs and warm',
'without reason. consider that the whale has nothing that can',
'provided with it. thereby, the whale commanders are enabled to',
'harpooneer shall remain on the whale till the whole tensing',
'flounders about, half on the whale and half in the',
'and then from between the whale and ship, which would',
'poor queequeg from between the whale and the ship--where he',
'in the tubs, and the whale not sounding very rapidly,',
'on the water, while the whale beyond also rose to',
'ahead, and let\'s get the whale alongside." "i think i',
'here hailed, to tow the whale on the larboard side,',
'but the ear of the whale is full as curious',
'vast a being as the whale should see the world',
'moreover you observe that the whale has no external nose;',
'so the tun of the whale contains by far the',
'grandeur. a nose to the whale would have been impertinent.',
'pitiable, and maddening sight. the whale was now going head',
'in the air. and the whale soon ceasing to sound,',
'the boats, and soon the whale broke water within two',
'more than sufferable anguish, the whale now spouting thick blood,',
'the ship drew nigh, the whale was transferred to her',
'and so low had the whale now settled that the',
'the true form of the whale was unknown to artists;',
'whaleman; at any rate the whale caught him, if he',
'story of hercules and the whale is considered to be',
'story of jonah and the whale in the preceding chapter.',
'who have opined that the whale mentioned in the book',
'jonah was swallowed by the whale in the mediterranean sea,',
```

'no other way for the whale to land the prophet',
"like a human being's, the whale can only live by",
'alike. now, why should the whale thus insist upon having',
'has been said that the whale only breathes through his',
'exclusion of water, therefore the whale has no voice; unless',
'then again, what has the whale to say? seldom have',
'the desert; even then, the whale always carries a small',
'the precise nature of the whale spout. it will not',
'the light buoyancy of the whale boat, and the elasticity',
'to me, that in the whale the sense of touch',
'when in maidenly gentleness the whale with a certain soft',
'pity it is that the whale does not possess this',
'fourth: stealing unawares upon the whale in the fancied security',
'in this chapter, between the whale and the elephant, so',
'of general bulk between the whale and the elephant is',
'the same respect to the whale that a dog does',
'wanting other motions of the whale in his general body,',
'on the part of the whale struck under such circumstances,',
'as, blind and deaf, the whale plunged forward, as if',
'moisture thrown off by the whale in his more quiet',
'laws and regulations of the whale fishery, of which the',
'had remained attached to the whale at the time of',
'that the examples of the whale and the lady were',
'meanwhile ruefully glancing from the whale to the stranger. but',
'his." in a word, the whale was seized and sold,',
'of the duke to the whale was a delegated one',
'for it. says plowdon, the whale so caught belongs to',
'it was plain that the whale alongside must be what',
'nose. nor indeed can the whale possibly be otherwise than',
'her own. now, in the whale ship, it is not',
'roared stubb; and so the whale was lost and pip',
'line; and hence, when the whale started to run, pip',
'turned upon him; and the whale was winged. in three',
'self-consuming misanthrope, once ignited, the whale supplies his own fue
l',
'from the back of the whale remains clinging to the',
'the digestive organs of the whale are so inscrutably constructed',
'is yours; only let the whale have another chance at',
'english were preceded in the whale fishery by the hollanders,',
'that the skeleton of the whale is by no means',
'from his mighty bulk the whale affords a most congenial',
'that the skeleton of the whale furnishes but little clue',
'was cast forth by the whale at the base of',
'this afric temple of the whale i leave you, reader,',
'believe these stories? no. the whale of to-day is as',
'that of all animals the whale alone should have degenerated.',
'these things, we account the whale immortal in his species,',
'my foster-brothers!" chapter 117 the whale watch. the four whales',
'the first sight of the whale myself,"--he said. "aye! ahab',
'heel. from this height the whale was now seen some',
'away--on each bright side, the whale shed off enticings. no',
'thrilled for an instant, the whale obliquely lying on his',
'in and out, as the whale dallied with the doomed',
'described. by this motion the whale must best and most',
'them he would take the whale head-and-head,--that is, pull straight',
'now for a moment the whale drew aside a little,',

```
'though!" when dusk descended, the whale was still in sight',
'him?" cried ahab; but the whale was not yet in',
'arm, ahab knew that the whale had sounded; but intending',
'strained planks; and as the whale swimming out from them,',
'during the past night, the whale had reeled the involutions',
'swim to feast on the whale or on ahab?--but pull',
'oblivious of its advance--as the whale sometimes will--and ahab was',
'of the sea-crashing boat, the whale wheeled round to present',
'beneath the settling ship, the whale ran quivering along its']
```

Build a lexicon of a word in a book of you choosing

Now, in the cell below, modify the code from above to search for a word in a book of your own choosing. Give your lexicon a meaningful name that is different from the ones above and comment your code so you understand what is happening each step of the way. But first lets have a look at the books you can choose from.

```
In [ ]: nltk.corpus.gutenberg.fileids()
```

```
Out[ ]: ['austen-emma.txt',
         'austen-persuasion.txt',
         'austen-sense.txt',
         'bible-kjv.txt',
         'blake-poems.txt',
         'bryant-stories.txt',
         'burgess-busterbrown.txt',
         'carroll-alice.txt',
         'chesterton-ball.txt',
         'chesterton-brown.txt',
         'chesterton-thursday.txt',
         'edgeworth-parents.txt',
         'melville-moby_dick.txt',
         'milton-paradise.txt',
         'shakespeare-caesar.txt',
         'shakespeare-hamlet.txt',
         'shakespeare-macbeth.txt',
         'whitman-leaves.txt']
```

```
In [ ]: # Add your code to build your own lexicon here, using the same method as abo
        macbeth = nltk.corpus.gutenberg.raw('shakespeare-macbeth.txt')

        macbeth = macbeth.lower()

        macbeth = macbeth.replace('\n', ' ')
        macbeth = macbeth.replace('\r', ' ')

        macbeth = macbeth.split()
```

Compare lengths of books

We can use the command `len` to find how many items there are in a list. E.g., to find the number of words in the list called `mobydick`, from earlier, we can write:

len(mobydick).

Use the starter code below to find out which book in the books included in `nltk` has the most words.

In [ ]:
```python
# One way to do it: Print all the titles and numbers of words
# starter code:

books = nltk.corpus.gutenberg.fileids()

for title in books:
    book = nltk.corpus.gutenberg.raw(title)
    len_book = len(book.split())
    print(f"{title}: {len_book} words")
```

```
austen-emma.txt: 158167 words
austen-persuasion.txt: 83308 words
austen-sense.txt: 118675 words
bible-kjv.txt: 821133 words
blake-poems.txt: 6845 words
bryant-stories.txt: 45988 words
burgess-busterbrown.txt: 15870 words
carroll-alice.txt: 26443 words
chesterton-ball.txt: 81598 words
chesterton-brown.txt: 71626 words
chesterton-thursday.txt: 57955 words
edgeworth-parents.txt: 166070 words
melville-moby_dick.txt: 212030 words
milton-paradise.txt: 79659 words
shakespeare-caesar.txt: 20459 words
shakespeare-hamlet.txt: 29605 words
shakespeare-macbeth.txt: 17741 words
whitman-leaves.txt: 122070 words
```

Optional more advanced way to do it, for those with python experience up for a challenge

In [ ]:
```python
# Another way to do it: Make a list of titles and a list of wordcounts, put
import pandas as pd


# starter code:
books = nltk.corpus.gutenberg.fileids()

titles = []
numwords = []

for title in books:
    book = nltk.corpus.gutenberg.raw(title)
    titles.append(title)
    numwords.append(len(book.split()))
    df = pd.DataFrame({'title': titles, 'numwords': numwords})
    df = df.sort_values(by='numwords', ascending=False)
```

```
print(df)
```

```
                    title   numwords
3            bible-kjv.txt     821133
12     melville-moby_dick.txt  212030
11       edgeworth-parents.txt  166070
0             austen-emma.txt   158167
17         whitman-leaves.txt   122070
2            austen-sense.txt   118675
1       austen-persuasion.txt    83308
8          chesterton-ball.txt    81598
13         milton-paradise.txt    79659
9        chesterton-brown.txt    71626
10   chesterton-thursday.txt     57955
5           bryant-stories.txt    45988
15     shakespeare-hamlet.txt    29605
7            carroll-alice.txt    26443
14     shakespeare-caesar.txt    20459
16   shakespeare-macbeth.txt    17741
6       burgess-busterbrown.txt  15870
4             blake-poems.txt     6845
```

Red excercises

Find the most frequent words in the book you choose earlier

`nltk` has a built-in function called `FreqDist` which counts up how many times each word in a text occurs. So, if you have a list called `words` which contains all the words in a book, you can find the frequencies of all of them by writing `freq = nltk.FreqDist(words)`. You can then get the ten most common words by writing `freq.most_common(10)` and so on.
What are the ten most common words in your book?

In [ ]:
```python
# starter code:

freq = nltk.FreqDist(macbeth)

print(f"The 10 most common words in Macbeth are: {freq.most_common(10)}")
```

```
The 10 most common words in Macbeth are: [('the', 645), ('and', 544), ('to',
383), ('of', 337), ('i', 314), ('a', 238), ('that', 220), ('my', 201), ('i
n', 195), ('is', 175)]
```

You might not feel like the ten most frequent words are of any real value in a potential project for analysis or comparison. This is due to the fact that the most frequent words in our language is what we call *stopwords*. These words like "a" and "the" are so common in English, that they don't really tell us much about the text.

That is why we often remove "stopwords", that is, a list of the most common words in

English, before e.g. counting frequencies.

There are several of these lists available, in English) as well as other languages, such as Danish.

Below is some starter code to remove stopwords. Use these snippets to see what the most common words in your book is after removing these most common words.

```
In [ ]:  # list of stopwords

stopwords = ["", "i", "me", "my", "myself", "we", "our", "ours", "ourselves"

# code to remove stopwords.
macbeth_nostop = [word for word in macbeth if word not in stopwords]

freq_nostop = nltk.FreqDist(macbeth_nostop)
print(f"The 10 most common words in Macbeth without stopwords are: {freq_nos
```

```
The 10 most common words in Macbeth without stopwords are: [('macb.', 137),
('haue', 119), ('thou', 85), ('enter', 74), ('shall', 67), ('vpon', 58), ('m
acd.', 58), ('thy', 56), ('hath', 52), ('yet', 52)]
```

Compare the pronouns used in books written by male and female authors

This could be an example of a project
One could imagine that authors draw most inspiration from there own life and therefore write about characters resembeling themselves.
Investigate whether male and femate authors of novels write more pronouns they identify with themselves.

Is there any evidence for this postulation in our data?

If you can, it would be neat to make a loop that goes through all books, and creates a dataframe that you afterwards can sort in to remove non-novels :)

```
In [ ]:  titles
```

```
Out[ 1]:  ['austen-emma.txt',
           'austen-persuasion.txt',
           'austen-sense.txt',
           'bible-kjv.txt',
           'blake-poems.txt',
           'bryant-stories.txt',
           'burgess-busterbrown.txt',
           'carroll-alice.txt',
           'chesterton-ball.txt',
           'chesterton-brown.txt',
           'chesterton-thursday.txt',
           'edgeworth-parents.txt',
           'melville-moby_dick.txt',
           'milton-paradise.txt',
           'shakespeare-caesar.txt',
           'shakespeare-hamlet.txt',
           'shakespeare-macbeth.txt',
           'whitman-leaves.txt']
```

In [ ]:
```
books
```

```
Out[ ]:  ['austen-emma.txt',
           'austen-persuasion.txt',
           'austen-sense.txt',
           'bible-kjv.txt',
           'blake-poems.txt',
           'bryant-stories.txt',
           'burgess-busterbrown.txt',
           'carroll-alice.txt',
           'chesterton-ball.txt',
           'chesterton-brown.txt',
           'chesterton-thursday.txt',
           'edgeworth-parents.txt',
           'melville-moby_dick.txt',
           'milton-paradise.txt',
           'shakespeare-caesar.txt',
           'shakespeare-hamlet.txt',
           'shakespeare-macbeth.txt',
           'whitman-leaves.txt']
```

In [ ]:
```
# Example of structure:

# Make a set of male and female pronouns (2 sets)
female_pronouns = {"she", "her", "hers", "herself"}
male_pronouns = {"he", "him", "his", "himself"}

author_gender = ["female", "female", "female", "unknown", "male", "male",
                 "male", "male", "male", "male", "male", "female", "male",
                 "male", "male", "male", "male", "male"]

# Make a dictionary of all titles in the corpus and the authors gender (1 di
book_dic = {"title": titles,
            "author_gender": author_gender}


# Make empty lists to store number of female pronouns and number of male pro
```

```
female_pronoun_count = []
male_pronoun_count = []
titles = []

# Make a loop that iterates over each book, splits the text to words, counts
for title in books:
    book = nltk.corpus.gutenberg.raw(title)
    titles.append(title)
    words = book.split()
    female_pronoun_occur = [word for word in words if word in female_pronoun
    female_pronoun_count.append(len(female_pronoun_occur))
    male_pronoun_occur = [word for word in words if word in male_pronouns]
    male_pronoun_count.append(len(male_pronoun_occur))




# Make dataframe, fill it with the lists above, and print it
df = pd.DataFrame({'title': titles, "author_gender": author_gender,
                   "female_pronouns": female_pronoun_count,
                   "male_pronouns": male_pronoun_count})

print(df)
```

```
                        title author_gender  female_pronouns  male_pronouns
0              austen-emma.txt        female             3719           2861
1         austen-persuasion.txt        female             1862           1637
2             austen-sense.txt        female             3521           2218
3                bible-kjv.txt       unknown             2461          21359
4              blake-poems.txt          male               39             87
5            bryant-stories.txt          male              606           1586
6       burgess-busterbrown.txt          male                3            793
7              carroll-alice.txt          male              730            212
8            chesterton-ball.txt          male              168           2209
9           chesterton-brown.txt          male              193           2172
10       chesterton-thursday.txt          male               13           1826
11         edgeworth-parents.txt        female             2408           4481
12        melville-moby_dick.txt          male              377           4596
13           milton-paradise.txt          male              338           1667
14         shakespeare-caesar.txt          male               12            372
15         shakespeare-hamlet.txt          male              111            529
16        shakespeare-macbeth.txt          male               47            255
17             whitman-leaves.txt          male              237            715
```

Discuss:

What is your findings and what is the limitations of this project