

Exercise 6

You are almost done. Nice job!

We have a couple more interesting problems for you before you go.

As always, run the setup code below before working on the questions.

```
In [ ]: # Run this cell to setup the exercise
import sys
from pathlib import Path
learntools_dir = Path().absolute().parents[1]
sys.path.append(str(learntools_dir))
from learntools.core import binder; binder.bind(globals())
from learntools.python.ex6 import *
print('Setup complete.')
```

Let's start with a string lightning round to warm up. What are the lengths of the strings below?

For each of the five strings below, predict what `len()` would return when passed that string. Use the variable `length` to record your answer, then run the cell to check whether you were right.

0a.

```
In [ ]: a = ""
length = ____
q0.a.check()
```

0b.

```
In [ ]: b = "it's ok"
length = ____
q0.b.check()
```

0c.

```
In [ ]: c = 'it\'s ok'
length = ____
q0.c.check()
```

0d.

```
In [ ]: d = ""hey""
length = ____
q0.d.check()
```

0e.

```
In [ ]: e = '\n'
length = ____
q0.e.check()
```

1.

There is a saying that "Data scientists spend 80% of their time cleaning data, and 20% of their time complaining about cleaning data." Let's see if you can write a function to help clean US zip code data. Given a string, it should return whether or not that string represents a valid zip code. For our purposes, a valid zip code is any string consisting of exactly 5 digits.

HINT: `str` has a method that will be useful here. Use `help(str)` to review a list of string methods.

```
In [ ]: def is_valid_zip(zip_code):
        """Returns whether the input string is a valid (5 digit) zip code
        """
        pass

# Check your answer
q1.check()
```

```
In [ ]: # q1.hint()
        # q1.solution()
```

2.

A researcher has gathered thousands of news articles. But she wants to focus her attention on articles including a specific word. Complete the function below to help her filter her list of articles.

Your function should meet the following criteria:

- Do not include documents where the keyword string shows up only as a part of a larger word. For example, if she were looking for the keyword "closed", you would not include the string "enclosed."
- She does not want you to distinguish upper case from lower case letters. So the phrase "Closed the case." would be included when the keyword is "closed"
- Do not let periods or commas affect what is matched. "It is closed." would be included when the keyword is "closed". But you can assume there are no other types of punctuation.

```
In [ ]: def word_search(doc_list, keyword):
        """
        Takes a list of documents (each document is a string) and a keyword.
        Returns list of the index values into the original list for all documents
        containing the keyword.

        Example:
        doc_list = ["The Learn Python Challenge Casino.", "They bought a car", "
        >>> word_search(doc_list, 'casino')
        >>> [0]
        """
        pass

# Check your answer
q2.check()
```

```
In [ ]: # q2.hint()
        # q2.solution()
```

3.

Now the researcher wants to supply multiple keywords to search for. Complete the function below to help her.

(You're encouraged to use the `word_search` function you just wrote when implementing this function. Reusing code in this way makes your programs more robust and readable - and it saves typing!)

```
In [ ]: def multi_word_search(doc_list, keywords):
        """
        Takes list of documents (each document is a string) and a list of keywords
        Returns a dictionary where each key is a keyword, and the value is a list
        (from doc_list) of the documents containing that keyword

        >>> doc_list = ["The Learn Python Challenge Casino.", "They bought a car", "
        >>> keywords = ['casino', 'they']
        >>> multi_word_search(doc_list, keywords)
        {'casino': [0, 1], 'they': [1]}
        """
        pass
```

```
# Check your answer  
q3.check()
```

```
In [ ]: # q3.solution()
```

Keep Going

You've learned a lot. But even the best programmers rely heavily on "libraries" of code from other programmers. You'll learn about that in [the last lesson](#).
