

Python Basics

Python was conceived by Guido Van Possum. You can use Python to build practically any piece of software.

1. Python is open source. It's free to use.
2. It's very easy to build packages in Python, which is code that you can share with other people to solve specific problems.

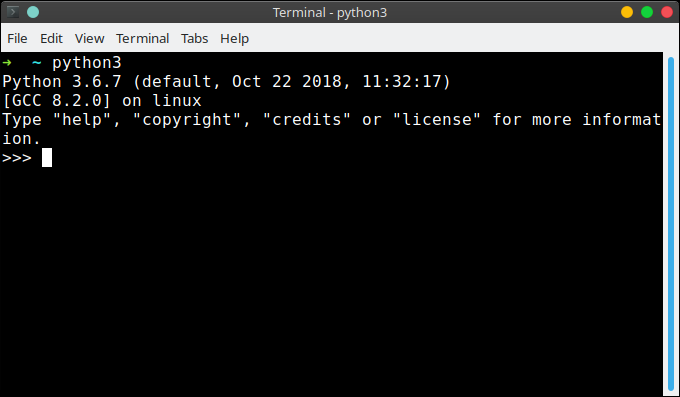
Throughout time, more and more of these packages specifically built for data science have been developed.

Suppose you want to make some fancy visualizations of your company's sales. There's a package for that.

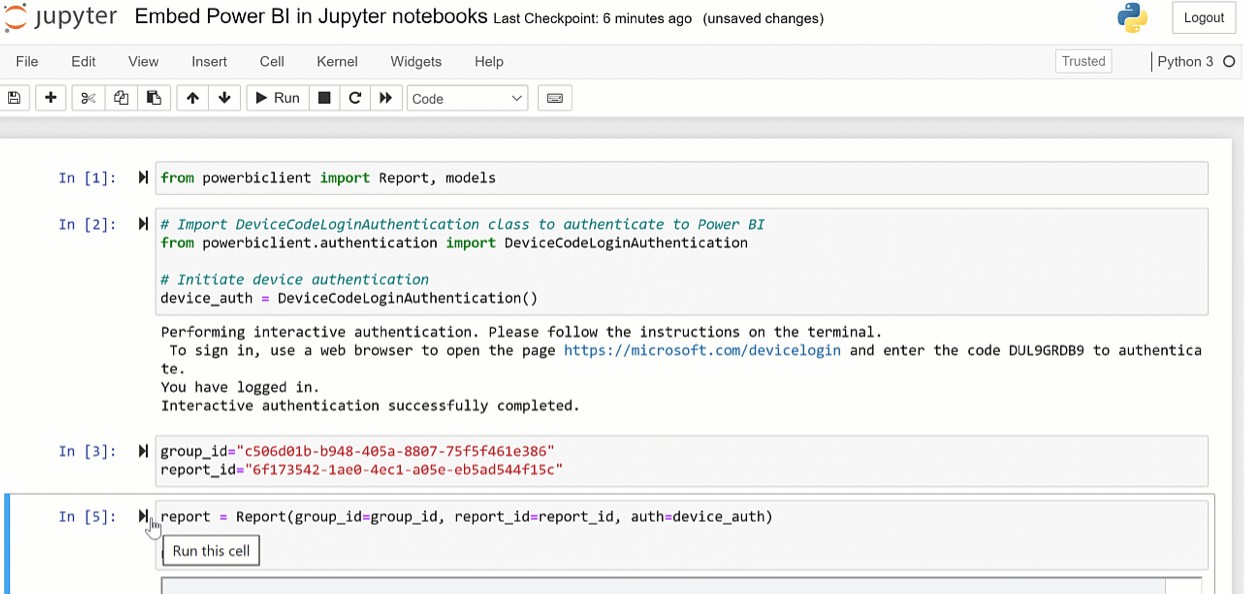
Or what about connecting to a database to analyze sensor measurements? There's also a package for that.

People often refer to Python as the swiss army knife of programming languages as you can do almost anything with it.

Python shell, a place where you can type Python code and immediately see the results.



The Python shell that's used here is actually not the original one; we're using IPython, short for Interactive Python, which is some kind of juiced up version of regular Python that'll be useful later on. IPython was created by Fernando Pérez and is part of the broader Jupyter ecosystem. Apart from interactively working with Python, you can also have Python run so called



The Python programming language has a wide range of syntactical constructions, standard library functions, and interactive development environment features.

Fortunately, you can ignore most of that; you just need to learn enough to write some handy little programs.

You will, however, have to learn some basic programming concepts before you can do anything. Like a wizard in training, you might think these concepts seem arcane and tedious, but with some knowledge and practice, you’ll be able to command your computer like a magic wand and perform incredible feats.

Python is the programming language of choice for data scientists. Although it wasn’t the first primary programming language, its popularity has grown throughout the years.

In 2016, it [overtook R on Kaggle](https://blog.kaggle.com/2017/01/05/your-year-on-kaggle-most-memorable-community-stats-from-2016/), the premier platform for data science competitions.

In 2017, it [overtook R on KDNuggets’s annual poll](https://www.kdnuggets.com/2017/08/python-overtakes-r-leader-analytics-data-science.html) of data scientists’ most-used tools.

In 2018, 66% of data scientists reported [using Python daily](https://www.kdnuggets.com/2018/05/poll-tools-analytics-data-science-machine-learning-results.html), making it the number one language for analytics professionals.

In 2021, it [overtook Java on the TIOBE index](https://content.techgig.com/python-retains-top-spot-as-most-popular-programming-language/articleshow/87581095.cms) and is now the most popular programming language.

What’s more, data science experts expect this trend to continue.

# Entering Expressions into the Interactive Shell

Enter **2 + 2** at the prompt to have Python do some simple math. The Mu window should now look like this:

>>> 2 + 2

4

>>>

In Python, 2 + 2 is called an *expression*, which is the most basic kind of programming instruction in the language. Expressions consist of *values* (such as 2)

and *operators* (such as +), and they can always *evaluate* (that is, reduce) down to a single value. That means you can use expressions anywhere in Python code that you could also use a value.

In the previous example, 2 + 2 is evaluated down to a single value, 4. A single value with no operators is also considered an expression, though it evaluates only to itself, as shown here:

>>> 2

2

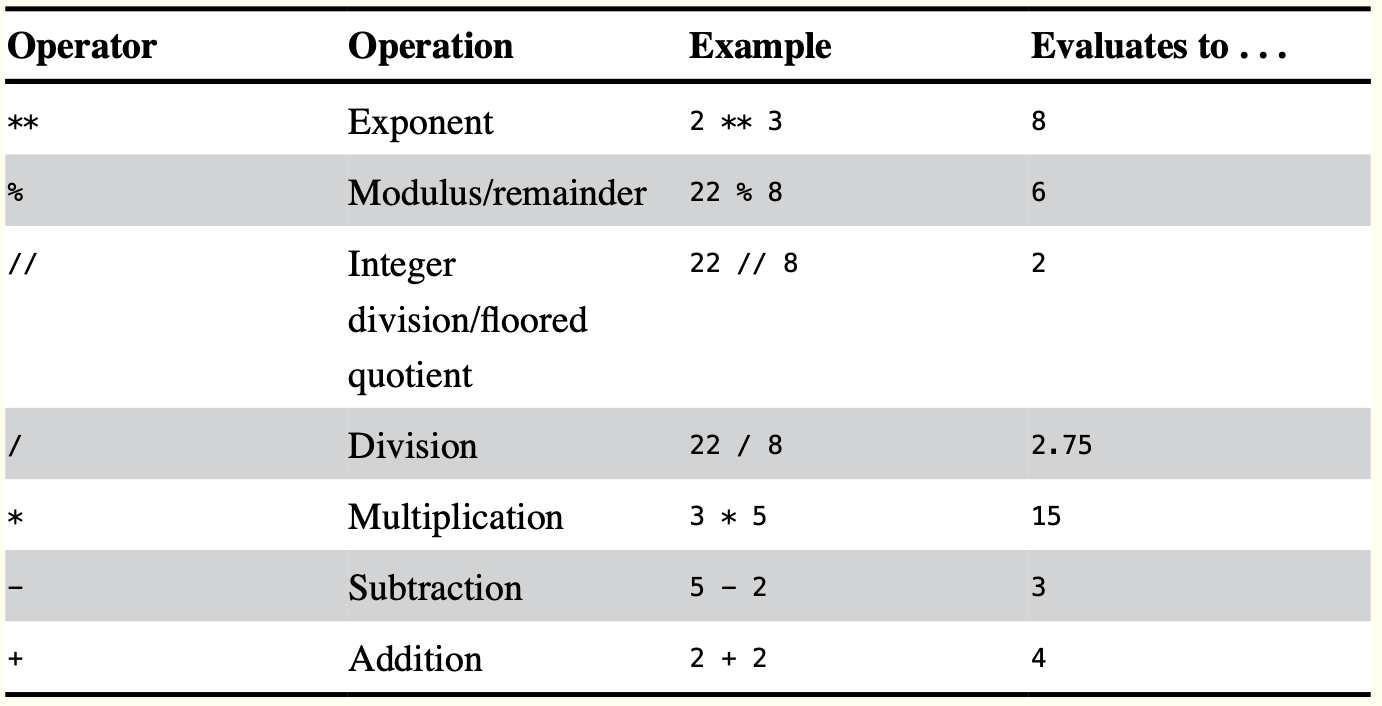
### ERRORS ARE OKAY!

Programs will crash if they contain code the computer can’t understand, which will cause Python to show an error message. An error message won’t break your computer, though, so don’t be afraid to make mistakes. A *crash* just means the program stopped running unexpectedly.

If you want to know more about an error, you can search for the exact error message text online for more information. You can also check out the resources

at [*https://nostarch.com/automatestuff2/*](https://nostarch.com/automatestuff2/)to see a list of common Python error messages and their meanings.

You can use plenty of other operators in Python expressions, too. For example,



**Table 1-1:** Math Operators from Highest to Lowest Precedence

The *order of operations* (also called *precedence*) of Python math operators is similar to



\*\*



%

that of mathematics. The

operator is evaluated first; the \* , / , // , and

operators

are evaluated next, from left to right; and the + and - operators are evaluated last (also from left to right). You can use parentheses to override the usual precedence if you need to. Whitespace in between the operators and values doesn’t matter for Python (except for the indentation at the beginning of the line), but a single space is convention. Enter the following expressions into the interactive shell:

>>> 2 + 3 \* 6

20

>>> (2 + 3) \* 6

30

>>> 48565878 \* 578453

28093077826734

>>> 2 \*\* 8

256

>>> 23 / 7

3.2857142857142856

>>> 23 // 7

3

>>> 23 % 7

2

>>> 2

4

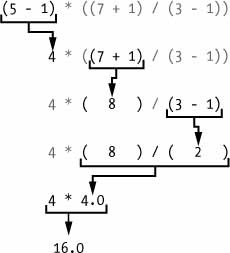
+

2

>>> (5 - 1) \* ((7 + 1) / (3 - 1))

16.0

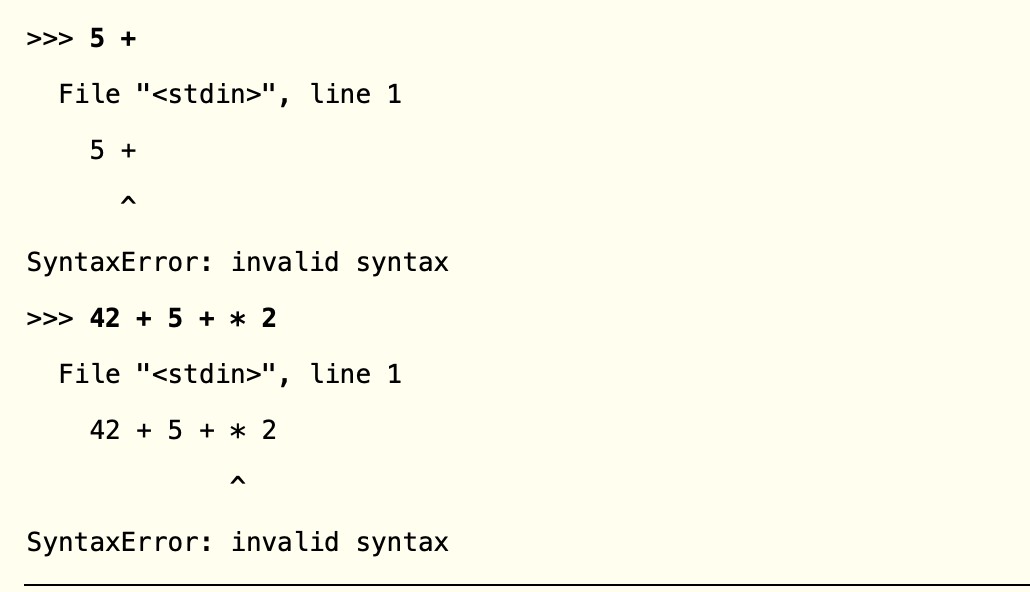
In each case, you as the programmer must enter the expression, but Python does the hard part of evaluating it down to a single value. Python will keep evaluating parts of the expression until it becomes a single value, as shown here:



These rules for putting operators and values together to form expressions are a fundamental part of Python as a programming language, just like the grammar rules that help us communicate. Here’s an example:

### This is a grammatically correct English sentence. This grammatically is sentence not English correct a.

The second line is difficult to parse because it doesn’t follow the rules of English. Similarly, if you enter a bad Python instruction, Python won’t be able to understand it and will display a SyntaxError error message, as shown here:

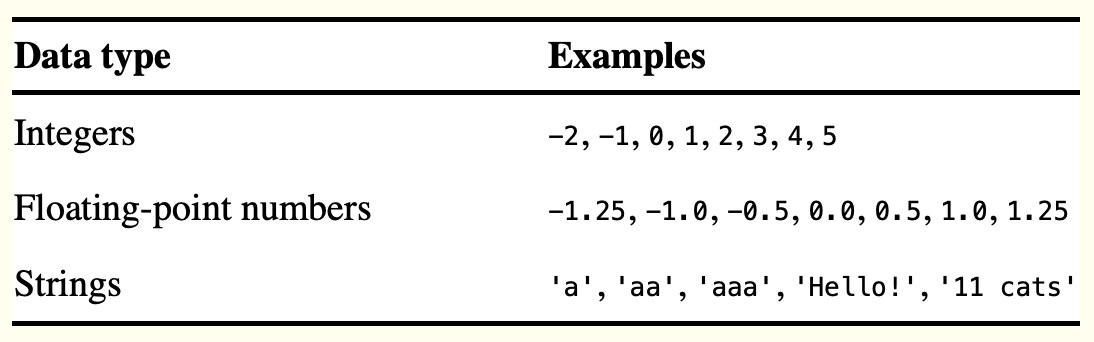


You can always test to see whether an instruction works by entering it into the interactive shell. Don’t worry about breaking the computer: the worst that could happen is that Python responds with an error message. Professional software developers get error messages while writing code all the time.

# The Integer, Floating-Point, and String Data Types

Remember that expressions are just values combined with operators, and they always evaluate down to a single value. A *data type* is a category for values, and every value belongs to exactly one data type. The most common data types in Python are listed

in below table. The values -2 and 30, for example, are said to be *integer* values. The integer (or *int*) data type indicates values that are whole numbers. Numbers with a decimal point, such as 3.14, are called *floating-point numbers* (or *floats*). Note that even though the value 42 is an integer, the value 42.0 would be a floating-point number.



**Table 1-2:** Common Data Types

Python programs can also have text values called *strings*, or *strs* (pronounced “stirs”). Always surround your string in single quote (') characters (as in 'Hello' or 'Goodbye cruel world!') so Python knows where the string begins and ends. You can even have a string with no characters in it, '', called a *blank string* or an *empty string*.

If you ever see the error message SyntaxError: EOL while scanning string literal, you probably forgot the final single quote character at the end of the string, such as in this example:

>>> 'Hello, world!

#SyntaxError: EOL while scanning string literal

# String Concatenation and Replication

The meaning of an operator may change based on the data types of the values next to it. For example, + is the addition operator when it operates on two integers or floating- point values. However, when + is used on two string values, it joins the strings as

the *string concatenation* operator. Enter the following into the interactive shell:

>>> 'Alice' + 'Bob' 'AliceBob'

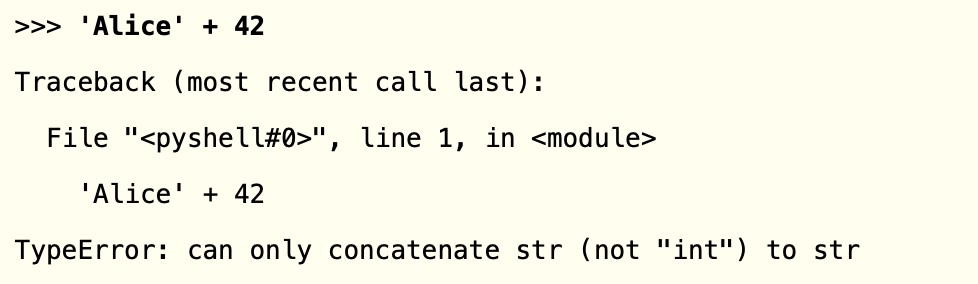
The expression evaluates down to a single, new string value that combines the text of

the two strings. However, if you try to use the operator on a string and an integer



+

value, Python will not know how to handle this, and it will display an error message.



The error message can only concatenate str (not "int") to str means that Python thought you were trying to concatenate an integer to the string 'Alice'. Your code will have to explicitly convert the integer to a string because Python cannot do this automatically.



\*



\*

The

operator multiplies two integer or floating-point values. But when the

operator

is used on one string value and one integer value, it becomes the *string replication* operator. Enter a string multiplied by a number into the interactive shell to see this in action.

>>> 'Alice' \* 5' 'AliceAliceAliceAliceAlice'

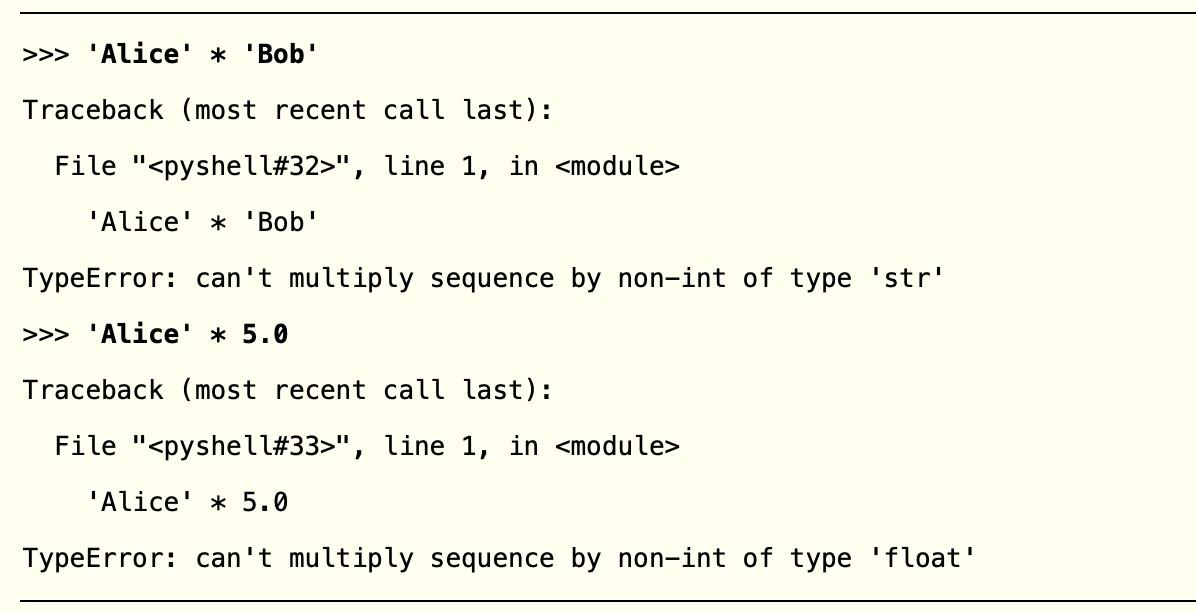
The expression evaluates down to a single string value that repeats the original string a number of times equal to the integer value. String replication is a useful trick, but it’s not used as often as string concatenation.

The operator can be used with only two numeric values (for multiplication), or one



\*

string value and one integer value (for string replication). Otherwise, Python will just display an error message, like the following:



It makes sense that Python wouldn’t understand these expressions: you can’t multiply two words, and it’s hard to replicate an arbitrary string a fractional number of times.

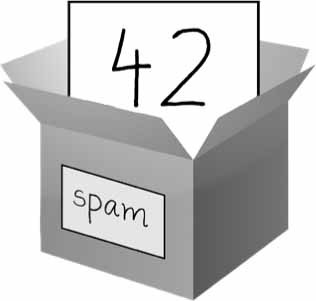
# Storing Values in Variables

A *variable* is like a box in the computer’s memory where you can store a single value. If you want to use the result of an evaluated expression later in your program, you can save it inside a variable.

## Assignment Statements

You’ll store values in variables with an *assignment statement*. An assignment statement consists of a variable name, an equal sign (called the *assignment operator*), and the value to be stored. If you enter the assignment statement spam = 42, then a variable named spam will have the integer value 42 stored in it.

Think of a variable as a labeled box that a value is placed in,



*Figure 1-1: spam = 42 is like telling the program, “The variable spam now has the integer value 42 in it.”*

For example, enter the following into the interactive shell:

➊ >>> spam = 40

>>> spam 40

>>> eggs = 2

➋ >>> spam + eggs 42

>>> spam + eggs + spam 82

* >>> spam = spam + 2

>>> spam 42

A variable is *initialized* (or created) the first time a value is stored in it ➊. After that, you can use it in expressions with other variables and values ➋. When a variable is assigned a new value ➌, the old value is forgotten, which is why spam evaluated to 42 instead of 40 at the end of the example. This is called *overwriting* the variable. Enter the following code into the interactive shell to try overwriting a string:

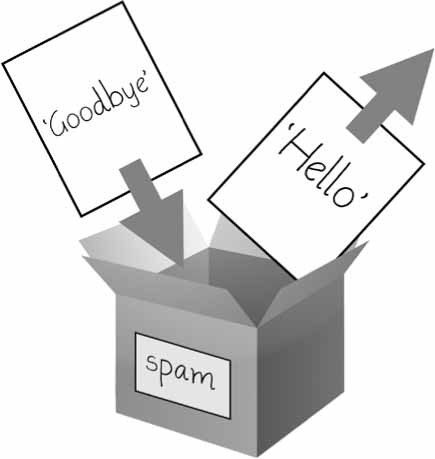
>>> spam = 'Hello'

>>> spam 'Hello'

>>> spam = 'Goodbye'

>>> spam 'Goodbye'

Just like the box in Figure 1-2, the spam variable in this example stores 'Hello' until you replace the string with 'Goodbye'.



*Figure 1-2: When a new value is assigned to a variable, the old one is forgotten.*

## Variable Names

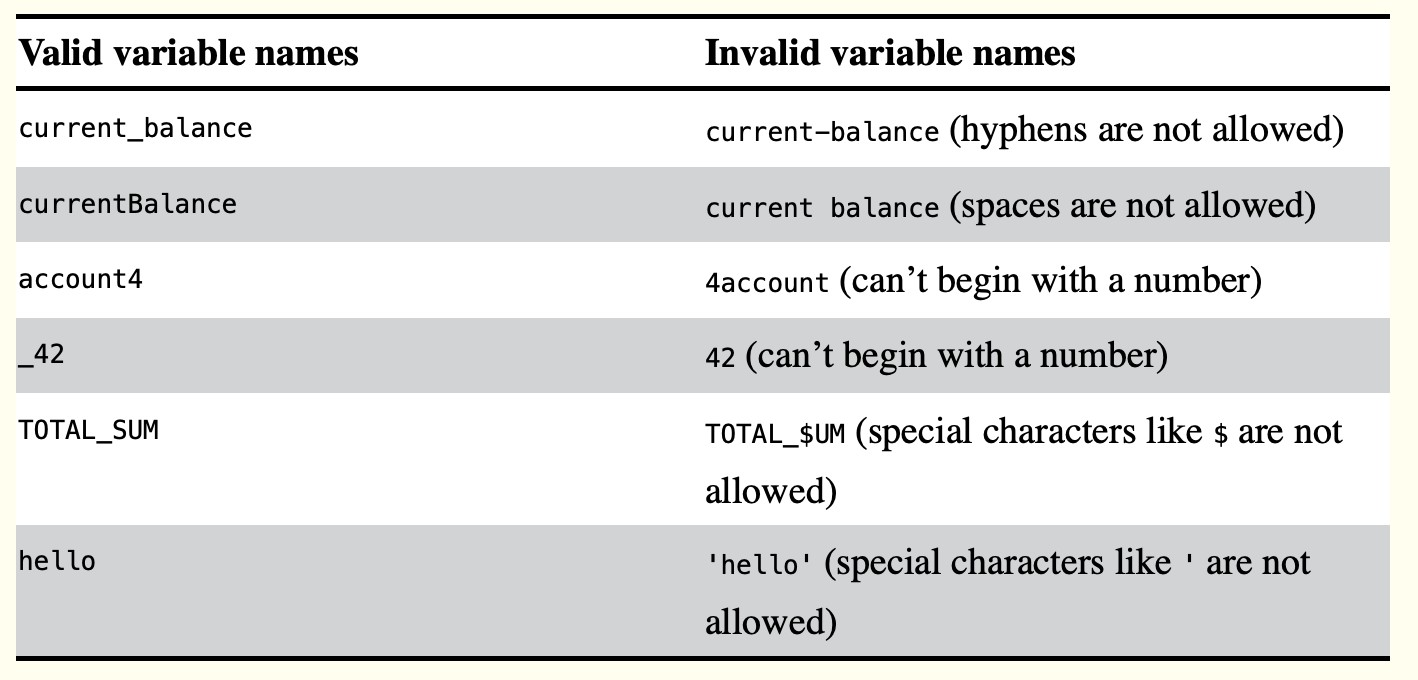
A good variable name describes the data it contains. Imagine that you moved to a new house and labeled all of your moving boxes as *Stuff*. You’d never find anything! Most of this book’s examples (and Python’s documentation) use generic variable names

like spam, eggs, and bacon, which come from the Monty Python “Spam” sketch. But in your programs, a descriptive name will help make your code more readable.

Though you can name your variables almost anything, Python does have some naming restrictions. Table 1-3 has examples of legal variable names. You can name a variable anything as long as it obeys the following three rules:

It can be only one word with no spaces.

It can use only letters, numbers, and the underscore (\_) character. It can’t begin with a number.



**Table 1-3:** Valid and Invalid Variable Names

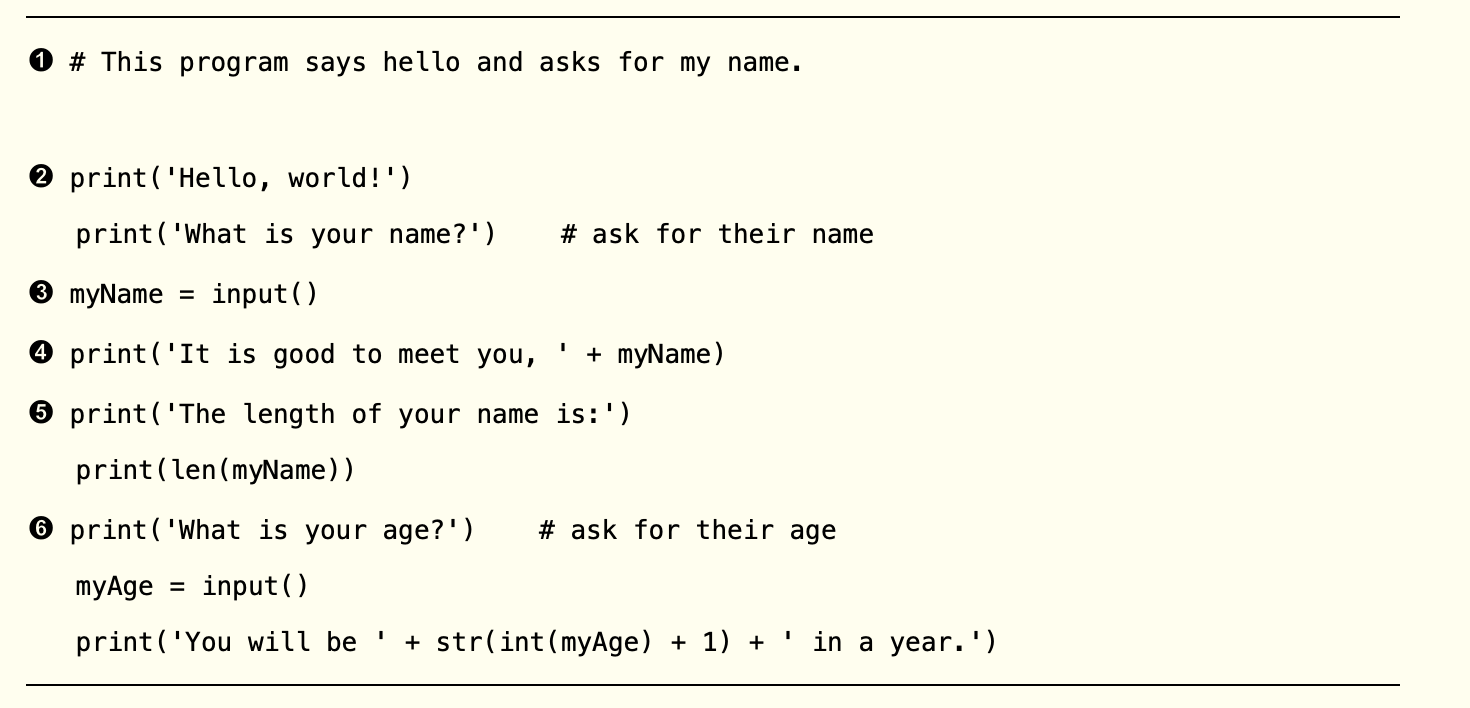
Variable names are case-sensitive, meaning that spam, SPAM, Spam, and sPaM are four different variables. Though Spam is a valid variable you can use in a program, it is a Python convention to start your variables with a lowercase letter.

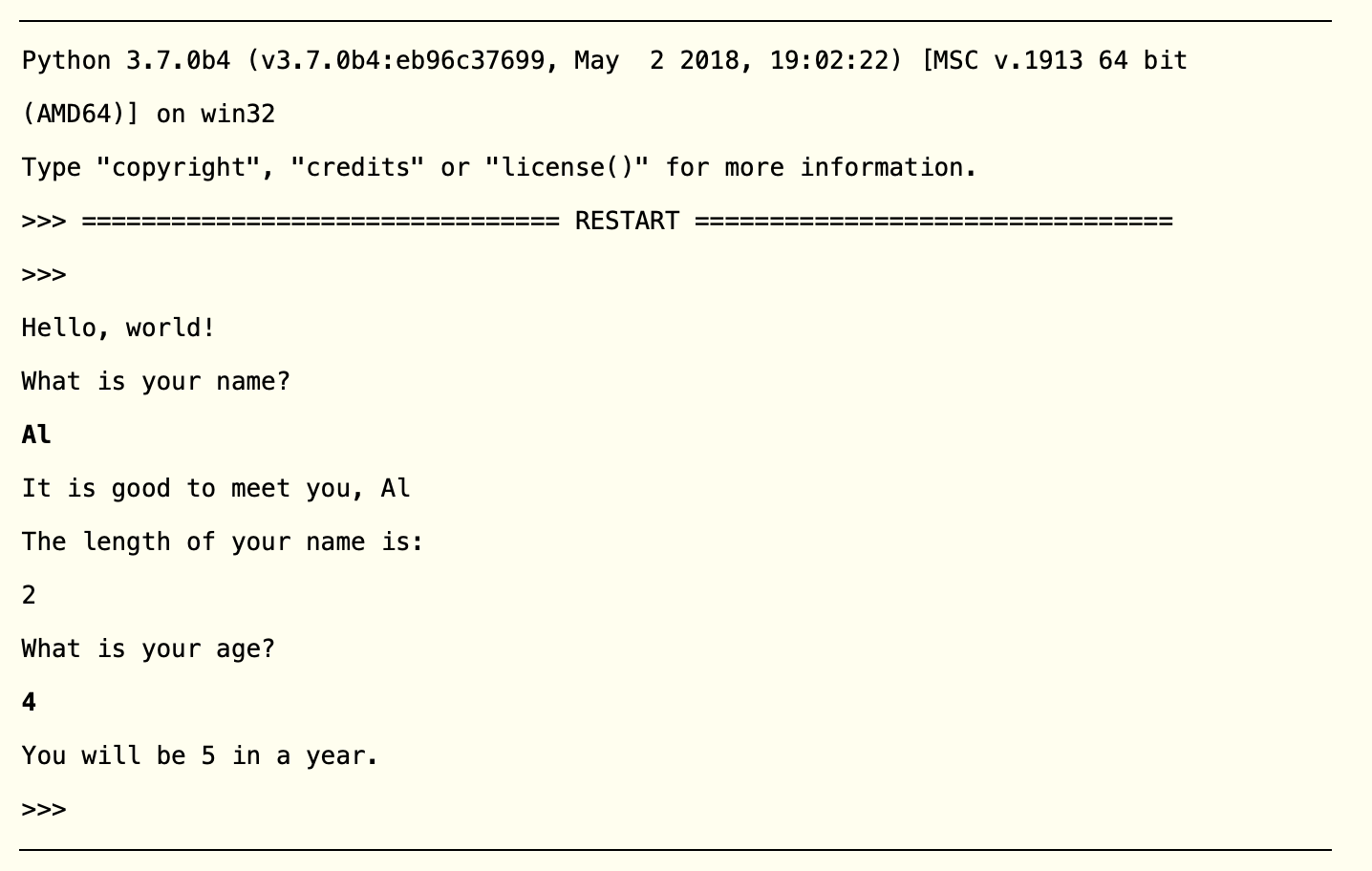
This book uses ***camelcase*** for variable names instead of underscores; that is, variables lookLikeThis instead of looking\_like\_this.

Some experienced programmers may point out that the official Python code style, PEP 8, says that underscores should be used. I unapologetically prefer camelcase and point to the “A Foolish Consistency Is the Hobgoblin of Little Minds” section in PEP 8 itself:

Consistency with the style guide is important. But most importantly: know when to be inconsistent—sometimes the style guide just doesn’t apply. When in doubt, use your best judgment.

# Your First Program





**Dissecting Your Program**

With your new program open in the file editor, let’s take a quick tour of the Python instructions it uses by looking at what each line of code does.

## Comments

The following line is called a *comment*.

➊ # This program says hello and asks for my name.

Python ignores comments, and you can use them to write notes or remind yourself what the code is trying to do. Any text for the rest of the line following a hash mark ( # ) is part of a comment.

Sometimes, programmers will put a in front of a line of code to temporarily remove it



#

while testing a program. This is called *commenting out* code, and it can be useful when you’re trying to figure out why a program isn’t working. You can remove the # later when you are ready to put the line back in.

Python also ignores the blank line after the comment. You can add as many blank lines to your program as you want. This can make your code easier to read, like paragraphs in a book.

## The print() Function

The print() function displays the string value inside its parentheses on the screen.

print('Hello, world!')

print('What is your name?') # ask for their name

The line means *“Print out the text in the string 'Hello, world!'.”*

print('Hello, world!')

When Python executes this line, you say that Python is *calling* the print() function and the string value is being *passed* to the function. A value that is passed to a function call is an *argument*. Notice that the quotes are not printed to the screen. They just mark where the string begins and ends; they are not part of the string value.

### NOTE

*You can also use this function to put a blank line on the screen; just call nothing in between the parentheses.*

*print()*

*with*

When you write a function name, the opening and closing parentheses at the end identify it as the name of a function. This is why in this pdf, you’ll see print() rather than print.

## The input() Function

The function waits for the user to type some text on the keyboard and

input()

press ENTER.

myName = input()

This function call evaluates to a string equal to the user’s text, and the line of code assigns the **myName** variable to this string value.

You can think of the function call as an expression that evaluates to whatever

input()

string the user typed in. If the user entered 'Al', then the expression would evaluate to

myName = 'Al'

If you call and see an error message, like NameError: name 'Al' is not defined,

input()

the problem is that you’re running the code with Python 2 instead of Python 3.

## Printing the User’s Name

The following call to actually contains the expression 'It is good to meet you, '

print()



+

myName between the parentheses.

print('It is good to meet you, ' + myName)

Remember that expressions can always evaluate to a single value. If 'Al' is the value stored in myName on line , then this expression evaluates to 'It is good to meet you, Al'. This single string value is then passed to print() , which prints it on the screen.

## The len() Function

You can pass the

len()

function a string value (or a variable containing a string), and

the function evaluates to the integer value of the number of characters in that string.

print('The length of your name is:') print(len(myName))

Enter the following into the interactive shell to try this:

>>> len('hello') 5

>>> len('My very energetic monster just scarfed nachos.') 46

>>> len('') 0

Just like those examples, evaluates to an integer. It is then passed

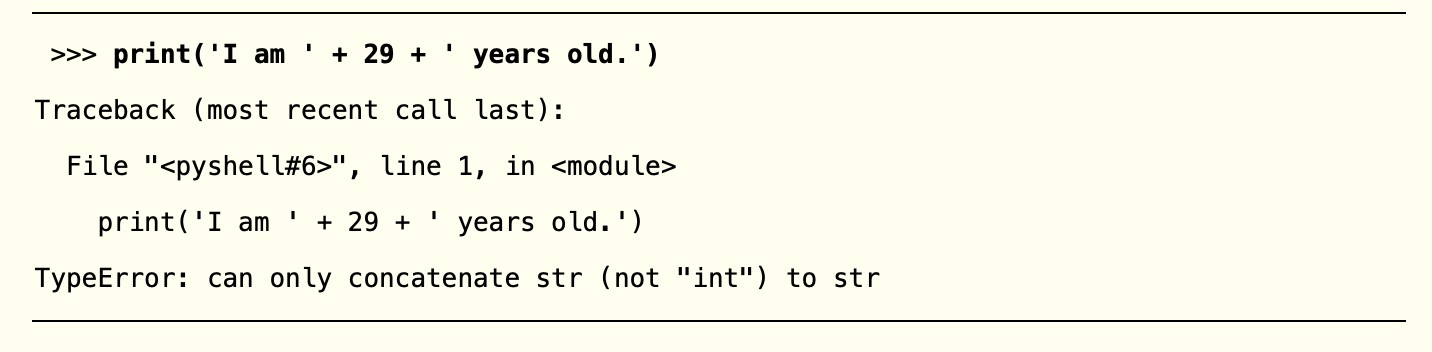
len(myName)

to to be displayed on the screen. The function allows you to pass it

print()

print()

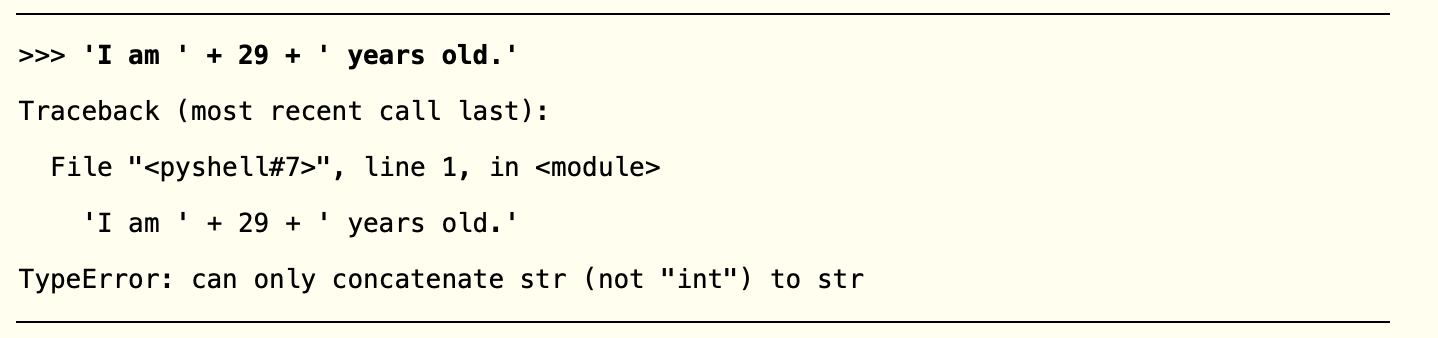
either integer values or string values, but notice the error that shows up when you type the following into the interactive shell:



The function isn’t causing that error, but rather it’s the expression you tried to

print()

pass to print() . You get the same error message if you type the expression into the interactive shell on its own.



Python gives an error because the operator can only be used to add two integers



+

together or concatenate two strings. You can’t add an integer to a string, because this is ungrammatical in Python. You can fix this by using a string version of the integer instead, as explained in the next section.

## The str(), int(), and float() Functions

If you want to concatenate an integer such as 29 with a string to pass to print() , you’ll

need to get the value '29', which is the string form of 29. The function can be

str()

passed an integer value and will evaluate to a string value version of the integer, as follows:

>>> str(29) '29'

>>> print('I am ' + str(29) + ' years old.') I am 29 years old.

Because evaluates to **'29'**, the expression

str(29)

'I am ' + str(29) + ' years

old.' evaluates to *'I am ' + '29' + ' years old.'*, which in turn evaluates to **'I am 29 years old.'.** This is the value that is passed to the print() function.

The str() , int() , and functions will evaluate to the string, integer, and floating-

float()

point forms of the value you pass, respectively. Try converting some values in the interactive shell with these functions and watch what happens.

>>> str(0) '0'

>>> str(-3.14)

'-3.14'

>>> int('42') 42

>>> int('-99')

-99

>>> int(1.25) 1

>>> int(1.99) 1

>>> float('3.14') 3.14

>>> float(10) 10.0

The previous examples call the str() , int() , and functions and pass them

float()

values of the other data types to obtain a string, integer, or floating-point form of those values.

The function is handy when you have an integer or float that you want to

str()

concatenate to a string. The function is also helpful if you have a number as a

int()

string value that you want to use in some mathematics. For example,

the Enter

input()

function always returns a string, even if the user enters a number.

into the interactive shell and enter **101** when it waits for your text.

**spam = input()**

>>> spam = input() 101

>>> spam '101'

The value stored inside spam isn’t the integer 101 but the string '101'. If you want to do math using the value in spam, use the int() function to get the integer form of spam and then store this as the new value in spam.

>>> spam = int(spam)

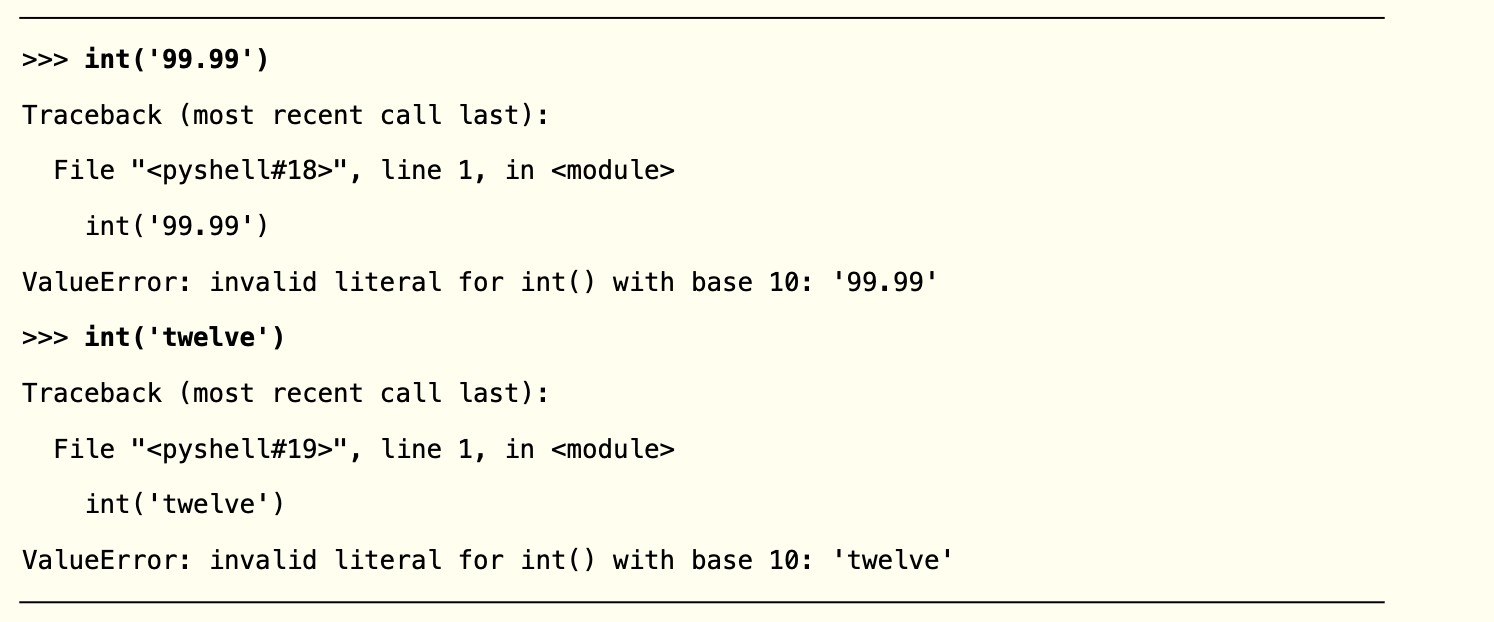
>>> spam 101

Now you should be able to treat the spam variable as an integer instead of a string.

>>> spam \* 10 / 5

202.0

Note that if you pass a value to int() that it cannot evaluate as an integer, Python will display an error message.



The function is also useful if you need to round a floating-point number down.

int()

>>> int(7.7) 7

>>> int(7.7) + 1 8

You used the

and

functions in the last three lines of your program to get a

value of the appropriate data type for the code.

int()

str()

print('What is your age?') # ask for their age myAge = input()

print('You will be ' + str(int(myAge) + 1) + ' in a year.')

### TEXT AND NUMBER EQUIVALENCE

Although the string value of a number is considered a completely different value from the integer or floating-point version, an integer can be equal to a floating point.

>>> 42 == '42'

False

>>> 42 == 42.0

True

>>> 42.0 == 0042.000

True

Python makes this distinction because strings are text, while integers and floats are both numbers.

The variable contains the value returned from input() . Because

myAge

the input() function always returns a string (even if the user typed in a number), you

can use the code to return an integer value of the string in **myAge**. This

int(myAge)

integer value is then added to 1 in the expression int(myAge) + 1 .

The result of this addition is passed to the function: str(int(myAge) + 1) . The

str()

string value returned is then concatenated with the strings ' You will be ' and to evaluate to one large string value. This large string is finally passed

' in a

year.'

to print() to be displayed on the screen.

Let’s say the user enters the string

for myAge . The string

is converted to an

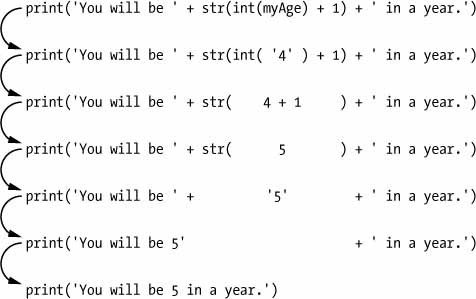
integer, so you can add one to it. The result is **5**. The function converts the result

'4'

'4'

str()

back to a string, so you can concatenate it with the second string, 'in a year.' , to create the final message. These evaluation steps would look something like the following:



# Summary

You can compute expressions with a calculator or enter string concatenations with a word processor. You can even do string replication easily by copying and pasting text. But expressions, and their component values—operators, variables, and function calls

—are the basic building blocks that make programs. Once you know how to handle

these elements, you will be able to instruct Python to operate on large amounts of data for you.

It is good to remember the different types of operators ( + , - , \* , / , // , % ,



+



\*

and

\*\*

for math operations, and

and

for string operations) and the three data

types (integers, floating-point numbers, and strings) introduced in this chapter.

I introduced a few different functions as well. The

and

functions handle

simple text output (to the screen) and input (from the keyboard). The len() function takes a string and evaluates to an int of the number of characters in the string.

print()

input()

The str() , int() , and functions will evaluate to the string, integer, or floating-

float()

point number form of the value they are passed.

In the next chapter, you’ll learn how to tell Python to make intelligent decisions about what code to run, what code to skip, and what code to repeat based on the values it has. This is known as *flow control*, and it allows you to write programs that make intelligent decisions.

# Practice Questions

1. Which of the following are operators, and which are values?

'hello'88.8/+5

1. Which of the following is a variable, and which is a string?

spam'spam'

1. Name three data types.
2. What is an expression made up of? What do all expressions do?
3. This chapter introduced assignment statements, like spam = 10. What is the difference between an expression and a statement?
4. What does the variable bacon contain after the following code runs?

bacon = 20bacon + 1

1. What should the following two expressions evaluate to?

'spam' + 'spamspam''spam' \* 3

1. Why is eggs a valid variable name while 100 is invalid?
2. What three functions can be used to get the integer, floating-point number, or string version of a value?
3. Why does this expression cause an error? How can you fix it?

'I have eaten ' + 99 + ' burritos.'

**Extra credit:** Search online for the Python documentation for the function. It will

len()

be on a web page titled “Built-in Functions.” Skim the list of other functions Python has,

look up what the function does, and experiment with it in the interactive shell.

round()