Writing Your first Python Code

Estaimted time needed: **25** minutes

Objectives

After compliting this lab you will be able to:

- Write basic code in Python
- Work with various types of data in python
- Convert the data from one type to another
- Use expressions and variables to perform operations

Table of Contents

Estimated time needed: **25 min**

Say "Hello" to the world in Python

When learning a new programming language, it is customary to start with an "hello world" example. As simple as it is, this one line of code will ensure that we know how to print a string in output and how to execute code within cells in a notebook.

[Tip]: To execute the Python code in the code cell below, click on the cell to select it and press `Shift` + `Enter`.

In [1]:
```
# Try your first Python output

print('Hello, Python!')
```

```
Hello, Python!
```

After executing the cell above, you should see that Python prints `Hello, Python!`. Congratulations on running your first Python code!

[Tip:] `print()` is a function. You passed the string `'Hello, Python!'` as an argument to instruct Python on what to print.

What version of Python are we using?

There are two popular versions of the Python programming language in use today: Python 2 and Python 3. The Python community has decided to move on from Python 2 to Python 3, and many popular libraries have announced that they will no longer support Python 2.

Since Python 3 is the future, in this course we will be using it exclusively. How do we know that our notebook is executed by a Python 3 runtime? We can look

in the top-right hand corner of this notebook and see "Python 3".

We can also ask directly Python and obtain a detailed answer. Try executing the following code:

In [2]:
```
# Check the Python Version

import sys
print(sys.version)
```

```
3.6.11 | packaged by conda-forge | (default, Jul 23 2020, 22:18:32)
[GCC 7.5.0]
```

[Tip:] `sys` is a built-in module that contains many system-specific parameters and functions, including the Python version in use. Before using it, we must explictly `import` it.

Writing comments in Python

In addition to writing code, note that it's always a good idea to add comments to your code. It will help others understand what you were trying to accomplish (the reason why you wrote a given snippet of code). Not only does this help **other people** understand your code, it can also serve as a reminder **to you** when you come back to it weeks or months later.

To write comments in Python, use the number symbol `#` before writing your comment. When you run your code, Python will ignore everything past the `#` on a given line.

In [3]:
```
# Practice on writing comments

print('Hello, Python!') # This line prints a string
# print('Hi')
```

```
Hello, Python!
```

After executing the cell above, you should notice that `This line prints a string` did not appear in the output, because it was a comment (and thus ignored by Python).

The second line was also not executed because `print('Hi')` was preceded by the number sign (`#`) as well! Since this isn't an explanatory comment from the programmer, but an actual line of code, we might say that the programmer *commented out* that second line of code.

Errors in Python

Everyone makes mistakes. For many types of mistakes, Python will tell you that you have made a mistake by giving you an error message. It is important to read error messages carefully to really understand where you made a mistake and how you may go about correcting it.

For example, if you spell `print` as `frint`, Python will display an error message. Give it a try:

In [4]:
```
# Print string as error message

frint("Hello, Python!")
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-4-313a1769a8a5> in <module>
      1 # Print string as error message
      2
----> 3 frint("Hello, Python!")

NameError: name 'frint' is not defined
```

The error message tells you:

1. where the error occurred (more useful in large notebook cells or scripts), and
2. what kind of error it was (NameError)

Here, Python attempted to run the function `frint`, but could not determine what `frint` is since it's not a built-in function and it has not been previously defined by us either.

You'll notice that if we make a different type of mistake, by forgetting to close the string, we'll obtain a different error (i.e., a `SyntaxError`). Try it below:

In [5]:
```
# Try to see build in error message

print("Hello, Python!)
```

```
  File "<ipython-input-5-63a21a726720>", line 3
    print("Hello, Python!)
                          ^
SyntaxError: EOL while scanning string literal
```

Does Python know about your error before it runs your code?

Python is what is called an *interpreted language*. Compiled languages examine your entire program at compile time, and are able to warn you about a whole class of errors prior to execution. In contrast, Python interprets your script line by line as it executes it. Python will stop executing the entire program when it encounters an error (unless the error is expected and handled by the programmer, a more advanced subject that we'll cover later on in this course).

Try to run the code in the cell below and see what happens:

In [6]:
```
# Print string and error to see the running order

print("This will be printed")
frint("This will cause an error")
print("This will NOT be printed")
```

```
This will be printed

---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-6-af59af1b345d> in <module>
      2
      3 print("This will be printed")
----> 4 frint("This will cause an error")
      5 print("This will NOT be printed")

NameError: name 'frint' is not defined
```

Exercise: Your First Program

Generations of programmers have started their coding careers by simply printing "Hello, world!". You will be following in their footsteps.

In the code cell below, use the `print()` function to print out the phrase: `Hello, world!`

In [7]:
```
print("Hello, world!")
```

```
Hello, world!
```

Double-click **here** for the solution.

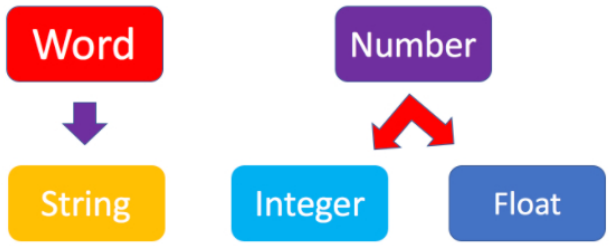
Now, let's enhance your code with a comment. In the code cell below, print out the phrase: `Hello, world!` and comment it with the phrase `Print the traditional hello world` all in one line of code.

In [8]:
```
#Print the traditional hello world
```

Double-click **here** for the solution.

Types of objects in Python

Python is an object-oriented language. There are many different types of objects in Python. Let's start with the most common object types: *strings*, *integers* and *floats*. Anytime you write words (text) in Python, you're using *character strings* (strings for short). The most common numbers, on the other hand, are *integers* (e.g. -1, 0, 100) and *floats*, which represent real numbers (e.g. 3.14, -42.0).



The following code cells contain some examples.

In [9]:
```
# Integer

11
```

Out[9]:
```
11
```

In [10]:
```
# Float

2.14
```

Out[10]:
```
2.14
```

In [11]:
```
# String

"Hello, Python 101!"
```

Out[11]:
```
'Hello, Python 101!'
```

You can get Python to tell you the type of an expression by using the built-in `type()` function. You'll notice that Python refers to integers as `int`, floats as `float`, and character strings as `str`.

In [12]:
```
# Type of 12
```

```
type(12)
```

Out[12]: int

In [13]:
```
# Type of 2.14

type(2.14)
```

Out[13]: float

In [14]:
```
# Type of "Hello, Python 101!"

type("Hello, Python 101!")
```

Out[14]: str

In the code cell below, use the `type()` function to check the object type of `12.0`.

In [15]:
```
type(12.0)
```

Out[15]: float

Double-click **here** for the solution.

Integers

Here are some examples of integers. Integers can be negative or positive numbers:

We can verify this is the case by using, you guessed it, the `type()` function:

In [16]:
```
# Print the type of -1

type(-1)
```

Out[16]: int

In [17]:
```
# Print the type of 4

type(4)
```

Out[17]: int

In [18]:
```
# Print the type of 0

type(0)
```

Out[18]: int

Floats

Floats represent real numbers; they are a superset of integer numbers but also include "numbers with decimals". There are some limitations when it comes to machines representing real numbers, but floating point numbers are a good representation in most cases. You can learn more about the specifics of floats for your runtime environment, by checking the value of `sys.float_info`. This will also tell you what's the largest and smallest number that can be represented with them.

Once again, can test some examples with the `type()` function:

In [19]:
```
# Print the type of 1.0

type(1.0) # Notice that 1 is an int, and 1.0 is a float
```

Out[19]: float

In [20]:
```
# Print the type of 0.5

type(0.5)
```

Out[20]: float

In [21]:
```
# Print the type of 0.56

type(0.56)
```

Out[21]: float

In [22]:
```
# System settings about float type

sys.float_info
```

Out[22]:
```
sys.float_info(max=1.7976931348623157e+308, max_exp=1024, max_10_exp=308, min=2.2250738585072014e-308, min_exp=-1021, min_10_exp=-307, dig=15, mant_dig=53, epsilon=2.220446049250313e-16, radix=2, rounds=1)
```

Converting from one object type to a different object type

You can change the type of the object in Python; this is called typecasting. For example, you can convert an *integer* into a *float* (e.g. 2 to 2.0).

Let's try it:

In [23]:
```
# Verify that this is an integer

type(2)
```

Out[23]:
```
int
```

Converting integers to floats

Let's cast integer 2 to float:

In [24]:
```
# Convert 2 to a float

float(2)
```

Out[24]:
```
2.0
```

In [25]:
```
# Convert integer 2 to a float and check its type

type(float(2))
```

Out[25]:
```
float
```

When we convert an integer into a float, we don't really change the value (i.e., the significand) of the number. However, if we cast a float into an integer, we could potentially lose some information. For example, if we cast the float 1.1 to integer we will get 1 and lose the decimal information (i.e., 0.1):

In [26]:
```
# Casting 1.1 to integer will result in loss of information

int(1.1)
```

Out[26]:
```
1
```

Converting from strings to integers or floats

Sometimes, we can have a string that contains a number within it. If this is the case, we can cast that string that represents a number into an integer using `int()`:

In [27]:
```
# Convert a string into an integer

int('1')
```

Out[27]:
```
1
```

But if you try to do so with a string that is not a perfect match for a number, you'll get an error. Try the following:

In [28]:
```
# Convert a string into an integer with error

int('1 or 2 people')
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-28-b78145d165c7> in <module>
      1 # Convert a string into an integer with error
      2
----> 3 int('1 or 2 people')

ValueError: invalid literal for int() with base 10: '1 or 2 people'
```

You can also convert strings containing floating point numbers into *float* objects:

In [29]:
```
# Convert the string "1.2" into a float

float('1.2')
```

Out[29]:
```
1.2
```

[Tip:] Note that strings can be represented with single quotes (`'1.2'`) or double quotes (`"1.2"`), but you can't mix both (e.g., `"1.2'`).

Converting numbers to strings

If we can convert strings to numbers, it is only natural to assume that we can convert numbers to strings, right?

In [30]:
```
# Convert an integer to a string

str(1)
```

Out[30]:
```
'1'
```

And there is no reason why we shouldn't be able to make floats into strings as well:

In [31]:
```
# Convert a float to a string

str(1.2)
```

Out[31]:
```
'1.2'
```

Boolean data type

Boolean is another important type in Python. An object of type *Boolean* can take on one of two values: `True` or `False`:

In [32]:
```
# Value true

True
```

Out[32]: True

Notice that the value `True` has an uppercase "T". The same is true for `False` (i.e. you must use the uppercase "F").

In [33]:
```
# Value false

False
```

Out[33]: False

When you ask Python to display the type of a boolean object it will show `bool` which stands for *boolean*:

In [34]:
```
# Type of True

type(True)
```

Out[34]: bool

In [35]:
```
# Type of False

type(False)
```

Out[35]: bool

We can cast boolean objects to other data types. If we cast a boolean with a value of `True` to an integer or float we will get a one. If we cast a boolean with a value of `False` to an integer or float we will get a zero. Similarly, if we cast a 1 to a Boolean, you get a `True`. And if we cast a 0 to a Boolean we will get a `False`. Let's give it a try:

In [36]:
```
# Convert True to int

int(True)
```

Out[36]: 1

In [37]:
```
# Convert 1 to boolean

bool(1)
```

Out[37]: True

In [38]:
```
# Convert 0 to boolean

bool(0)
```

Out[38]: False

In [39]:
```
# Convert True to float

float(True)
```

Out[39]: 1.0

Exercise: Types

What is the data type of the result of: `6 / 2`?

In [40]:
```
# Write your code below. Don't forget to press Shift+Enter to execute the cell
```

Double-click **here** for the solution.

What is the type of the result of: `6 // 2`? (Note the double slash `//`.)

In [41]:
```
# Write your code below. Don't forget to press Shift+Enter to execute the cell
```

Double-click **here** for the solution.

Expression and Variables

Expressions

Expressions in Python can include operations among compatible types (e.g., integers and floats). For example, basic arithmetic operations like adding multiple numbers:

In [42]:
```
# Addition operation expression

43 + 60 + 16 + 41
```

Out[42]: 160

We can perform subtraction operations using the minus operator. In this case the result is a negative number:

In [43]:
```
# Subtraction operation expression

50 - 60
```

Out[43]: -10

We can do multiplication using an asterisk:

In [44]:
```
# Multiplication operation expression

5 * 5
```

Out[44]: 25

We can also perform division with the forward slash:

In [45]:
```
# Division operation expression

25 / 5
```

Out[45]: 5.0

In [46]:
```
# Division operation expression

25 / 6
```

Out[46]: 4.166666666666667

As seen in the quiz above, we can use the double slash for integer division, where the result is rounded to the nearest integer:

In [47]:
```
# Integer division operation expression

25 // 5
```

Out[47]: 5

In [48]:
```
# Integer division operation expression

25 // 6
```

Out[48]: 4

Exercise: Expression

Let's write an expression that calculates how many hours there are in 160 minutes:

In [49]:
```
160/60
```

Out[49]: 2.6666666666666665

Double-click **here** for the solution.

Python follows well accepted mathematical conventions when evaluating mathematical expressions. In the following example, Python adds 30 to the result of the multiplication (i.e., 120).

In [50]:
```
# Mathematical expression

30 + 2 * 60
```

Out[50]: 150

And just like mathematics, expressions enclosed in parentheses have priority. So the following multiplies 32 by 60.

In [51]:
```
# Mathematical expression

(30 + 2) * 60
```

Out[51]: 1920

Variables

Just like with most programming languages, we can store values in *variables*, so we can use them later on. For example:

In [52]:
```
# Store value into variable

x = 43 + 60 + 16 + 41
```

To see the value of `x` in a Notebook, we can simply place it on the last line of a cell:

In [53]:
```
# Print out the value in variable

x
```

Out[53]: 160

We can also perform operations on `x` and save the result to a new variable:

In [54]:
```
# Use another variable to store the result of the operation between variable and value

y = x / 60
y
```

Out[54]: 2.6666666666666665

If we save a value to an existing variable, the new value will overwrite the previous value:

In [55]:
```
# Overwrite variable with new value

x = x / 60
x
```

Out[55]: 2.6666666666666665

It's a good practice to use meaningful variable names, so you and others can read the code and understand it more easily:

In [56]:
```
# Name the variables meaningfully

total_min = 43 + 42 + 57 # Total length of albums in minutes
total_min
```

Out[56]: 142

In [57]:
```
# Name the variables meaningfully

total_hours = total_min / 60 # Total length of albums in hours
total_hours
```

Out[57]: 2.3666666666666667

In the cells above we added the length of three albums in minutes and stored it in `total_min`. We then divided it by 60 to calculate total length `total_hours` in hours. You can also do it all at once in a single expression, as long as you use parenthesis to add the albums length before you divide, as shown below.

In [58]:
```
# Complicate expression

total_hours = (43 + 42 + 57) / 60  # Total hours in a single expression
total_hours
```

Out[58]: 2.3666666666666667

If you'd rather have total hours as an integer, you can of course replace the floating point division with integer division (i.e., `//`).

Exercise: Expression and Variables in Python

What is the value of `x` where `x = 3 + 2 * 2`

In [59]:
```
x = 3+2*2
x
```

Out[59]: 7

Double-click **here** for the solution.

What is the value of `y` where `y = (3 + 2) * 2`?

In [60]:
```
y = (3+2)*2
y
```

Out[60]: 10

Double-click **here** for the solution.

What is the value of `z` where `z = x + y`?

In [61]:
```
z = x+y
```

Double-click **here** for the solution.

The last exercise!

Congratulations, you have completed your first lesson and hands-on lab in Python. However, there is one more thing you need to do. The Data Science community encourages sharing work. The best way to share and showcase your work is to share it on GitHub. By sharing your notebook on GitHub you are not only building your reputation with fellow data scientists, but you can also show it off when applying for a job. Even though this was your first piece of work, it is never too early to start building good habits. So, please read and follow this article to learn how to share your work.

Author

Author

Joseph Santarcangelo

Other contributors

Mavis Zhou

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2020-08-26	2.0	Lavanya	Moved lab to course repo in GitLab

String Operations

Estaimted time needed: **15** minutes

Objectives

After compliting this lab you will be able to:

- Work with Strings
- Perform operations on String
- Manipulate Strings using indexing and escape sequences

Table of Contents

Estimated time needed: **15 min**

What are Strings?

The following example shows a string contained within 2 quotation marks:

In [1]:
```
# Use quotation marks for defining string

"Michael Jackson"
```

Out[1]: 'Michael Jackson'

We can also use single quotation marks:

In [2]:
```
# Use single quotation marks for defining string

'Michael Jackson'
```

Out[2]: 'Michael Jackson'

A string can be a combination of spaces and digits:

In [3]:
```
# Digitals and spaces in string

'1 2 3 4 5 6 '
```

Out[3]: '1 2 3 4 5 6 '

A string can also be a combination of special characters :

In [4]:
```
# Special characters in string

'@#2_#]&*^%$'
```

Out[4]: '@#2_#]&*^%$'

We can print our string using the print statement:

In [5]:
```
# Print the string

print("hello!")
```

```
hello!
```

We can bind or assign a string to another variable:

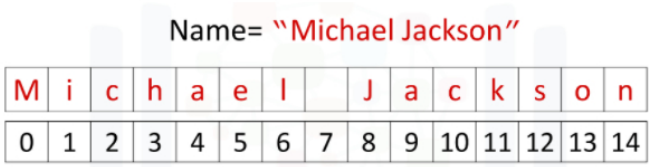
In [6]:
```
# Assign string to variable
```

```
Name = "Michael Jackson"
Name
```

Out[6]: 'Michael Jackson'

Indexing

It is helpful to think of a string as an ordered sequence. Each element in the sequence can be accessed using an index represented by the array of numbers:



The first index can be accessed as follows:

> [Tip]: Because indexing starts at 0, it means the first index is on the index 0.

In [7]:
```
# Print the first element in the string

print(Name[0])
```

```
M
```

We can access index 6:

In [8]:
```
# Print the element on index 6 in the string

print(Name[6])
```

```
l
```

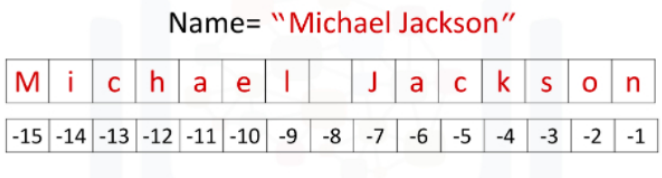
Moreover, we can access the 13th index:

In [9]:
```
# Print the element on the 13th index in the string

print(Name[13])
```

```
o
```

Negative Indexing

We can also use negative indexing with strings:



Negative index can help us to count the element from the end of the string.

The last element is given by the index -1:

In [10]:
```
# Print the last element in the string

print(Name[-1])
```

```
n
```

The first element can be obtained by index -15:

In [11]:
```
# Print the first element in the string

print(Name[-15])
```

```
M
```

We can find the number of characters in a string by using `len`, short for length:

In [12]:
```
# Find the length of string

len("Michael Jackson")
```

Out[12]: 15

Slicing

We can obtain multiple characters from a string using slicing, we can obtain the 0 to 4th and 8th to the 12th element:

[Tip]: When taking the slice, the first number means the index (start at 0), and the second number means the length from the index to the last element you want (start at 1)

In [13]:
```
# Take the slice on variable Name with only index 0 to index 3

Name[0:4]
```

Out[13]: 'Mich'

In [14]:
```
# Take the slice on variable Name with only index 8 to index 11

Name[8:12]
```

Out[14]: 'Jack'

Stride

We can also input a stride value as follows, with the '2' indicating that we are selecting every second variable:

In [17]:
```
# Get every second element. The elments on index 1, 3, 5 ...

Name[::2]
Name[::-1]
```

Out[17]: 'noskcaJ leahciM'

We can also incorporate slicing with the stride. In this case, we select the first five elements and then use the stride:

In [18]:
```
# Get every second element in the range from index 0 to index 4

Name[0:5:2]
```

Out[18]: 'Mca'

Concatenate Strings

We can concatenate or combine strings by using the addition symbols, and the result is a new string that is a combination of both:

In [19]:
```
# Concatenate two strings

Statement = Name + "is the best"
Statement
```

Out[19]: 'Michael Jacksonis the best'

To replicate values of a string we simply multiply the string by the number of times we would like to replicate it. In this case, the number is three. The result is a new string, and this new string consists of three copies of the original string:

In [20]:
```
# Print the string for 3 times

3 * "Michael Jackson"
```

```
Out[20]: 'Michael JacksonMichael JacksonMichael Jackson'
```

You can create a new string by setting it to the original variable. Concatenated with a new string, the result is a new string that changes from Michael Jackson to "Michael Jackson is the best".

```
In [21]: # Concatenate strings

Name = "Michael Jackson"
Name = Name + " is the best"
Name
```

```
Out[21]: 'Michael Jackson is the best'
```

Escape Sequences

Back slashes represent the beginning of escape sequences. Escape sequences represent strings that may be difficult to input. For example, back slash "n" represents a new line. The output is given by a new line after the back slash "n" is encountered:

```
In [22]: # New line escape sequence

print(" Michael Jackson \n is the best" )
```

```
 Michael Jackson
 is the best
```

Similarly, back slash "t" represents a tab:

```
In [23]: # Tab escape sequence

print(" Michael Jackson \t is the best" )
```

```
 Michael Jackson         is the best
```

If you want to place a back slash in your string, use a double back slash:

```
In [24]: # Include back slash in string

print(" Michael Jackson \\ is the best" )
```

```
 Michael Jackson \ is the best
```

We can also place an "r" before the string to display the backslash:

```
In [25]: # r will tell python that string will be display as raw string

print(r" Michael Jackson \ is the best" )
```

```
 Michael Jackson \ is the best
```

String Operations

There are many string operation methods in Python that can be used to manipulate the data. We are going to use some basic string operations on the data.

Let's try with the method `upper`; this method converts lower case characters to upper case characters:

```
In [26]: # Convert all the characters in string to upper case

A = "Thriller is the sixth studio album"
print("before upper:", A)
B = A.upper()
print("After upper:", B)
```

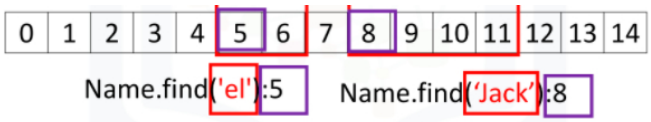
```
before upper: Thriller is the sixth studio album
After upper: THRILLER IS THE SIXTH STUDIO ALBUM
```

The method `replace` replaces a segment of the string, i.e. a substring with a new string. We input the part of the string we would like to change. The second argument is what we would like to exchange the segment with, and the result is a new string with the segment changed:

```
In [27]: # Replace the old substring with the new target substring is the segment has been found in the string

A = "Michael Jackson is the best"
B = A.replace('Michael', 'Janet')
B
```

```
Out[27]: 'Janet Jackson is the best'
```

The method `find` finds a sub-string. The argument is the substring you would like to find, and the output is the first index of the sequence. We can find the sub-string `jack` or `el .`



In [28]:
```
# Find the substring in the string. Only the index of the first elment of substring in string will be the output

Name = "Michael Jackson"
Name.find('el')
```

Out[28]: 5

In [29]:
```
# Find the substring in the string.

Name.find('Jack')
```

Out[29]: 8

If the sub-string is not in the string then the output is a negative one. For example, the string 'Jasdfasdasdf' is not a substring:

In [30]:
```
# If cannot find the substring in the string

Name.find('Jasdfasdasdf')
```

Out[30]: -1

Quiz on Strings

What is the value of the variable `A` after the following code is executed?

In [32]:
```
# Write your code below and press Shift+Enter to execute

A = "1"
A
```

Out[32]: '1'

Double-click **here** for the solution.

What is the value of the variable `B` after the following code is executed?

In [34]:
```
# Write your code below and press Shift+Enter to execute

B = "2"
B
```

Out[34]: '2'

Double-click **here** for the solution.

What is the value of the variable `C` after the following code is executed?

In [36]:
```
# Write your code below and press Shift+Enter to execute

C = A + B
C
```

Out[36]: '12'

Double-click **here** for the solution.

Consider the variable `D` use slicing to print out the first three elements:

In [37]:
```
# Write your code below and press Shift+Enter to execute

D = "ABCDEFG"
D[0:4]
```

Out[37]: 'ABCD'

Double-click **here** for the solution.

Use a stride value of 2 to print out every second character of the string `E` :

In [38]:
```
# Write your code below and press Shift+Enter to execute

E = 'clocrkr1e1c1t'
E[::2]
```

Out[38]: 'correct'

Double-click **here** for the solution.

Print out a backslash:

In [40]:
```
# Write your code below and press Shift+Enter to execute
print("\\")
```

```
\
```

Double-click **here** for the solution.

Convert the variable `F` to uppercase:

In [41]:
```
# Write your code below and press Shift+Enter to execute

F = "You are wrong"
F.upper()
```

Out[41]:
```
'YOU ARE WRONG'
```

Double-click **here** for the solution.

Consider the variable `G`, and find the first index of the sub-string `snow`:

In [42]:
```
# Write your code below and press Shift+Enter to execute

G = "Mary had a little lamb Little lamb, little lamb Mary had a little lamb \
Its fleece was white as snow And everywhere that Mary went Mary went, Mary went \
Everywhere that Mary went The lamb was sure to go"
G.find("snow")
```

Out[42]:
```
95
```

Double-click **here** for the solution.

In the variable `G`, replace the sub-string `Mary` with `Bob`:

In [43]:
```
# Write your code below and press Shift+Enter to execute
G.replace("Mary", "Bob")
```

Out[43]:
```
'Bob had a little lamb Little lamb, little lamb Bob had a little lamb Its fleece was white as snow And everywhere that Bob went Bob went, Bob went Everywhere that Bob went The lamb was sure to go'
```

Double-click **here** for the solution.

The last exercise!

Congratulations, you have completed your first lesson and hands-on lab in Python. However, there is one more thing you need to do. The Data Science community encourages sharing work. The best way to share and showcase your work is to share it on GitHub. By sharing your notebook on GitHub you are not only building your reputation with fellow data scientists, but you can also show it off when applying for a job. Even though this was your first piece of work, it is never too early to start building good habits. So, please read and follow this article to learn how to share your work.

Author

Joseph Santarcangelo

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2020-08-26	2.0	Lavanya	Moved lab to course repo in GitLab

In [ ]:
```

```

Lists in Python

Estimated time needed: **15** minutes

Objectives

After completing this lab you will be able to:

- Perform list operations in Python, including indexing, list manipulation and copy/clone list.

Table of Contents

About the Dataset

Imagine you received album recommendations from your friends and compiled all of the recommandations into a table, with specific information about each album.

The table has one row for each movie and several columns:

- **artist** - Name of the artist
- **album** - Name of the album
- **released_year** - Year the album was released
- **length_min_sec** - Length of the album (hours,minutes,seconds)
- **genre** - Genre of the album
- **music_recording_sales_millions** - Music recording sales (millions in USD) on SONG://DATABASE
- **claimed_sales_millions** - Album's claimed sales (millions in USD) on SONG://DATABASE
- **date_released** - Date on which the album was released
- **soundtrack** - Indicates if the album is the movie soundtrack (Y) or (N)
- **rating_of_friends** - Indicates the rating from your friends from 1 to 10

The dataset can be seen below:

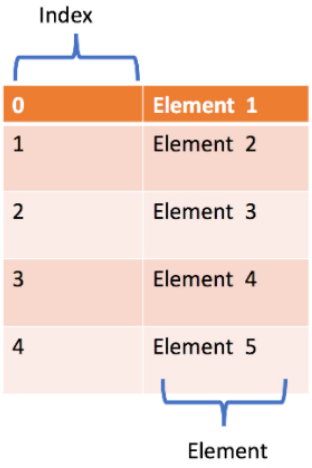
Artist	Album	Released	Length	Genre	Music recording sales (millions)	Claimed sales (millions)	Released	Soundtrack	Rating (friends)
Michael Jackson	Thriller	1982	00:42:19	Pop, rock, R&B	46	65	30-Nov-82		10.0
AC/DC	Back in Black	1980	00:42:11	Hard rock	26.1	50	25-Jul-80		8.5
Pink Floyd	The Dark Side of the Moon	1973	00:42:49	Progressive rock	24.2	45	01-Mar-73		9.5
Whitney Houston	The Bodyguard	1992	00:57:44	Soundtrack/R&B, soul, pop	26.1	50	25-Jul-80	Y	7.0
Meat Loaf	Bat Out of Hell	1977	00:46:33	Hard rock, progressive rock	20.6	43	21-Oct-77		7.0
Eagles	Their Greatest Hits (1971-1975)	1976	00:43:08	Rock, soft rock, folk rock	32.2	42	17-Feb-76		9.5
Bee Gees	Saturday Night Fever	1977	1:15:54	Disco	20.6	40	15-Nov-77	Y	9.0
Fleetwood Mac	Rumours	1977	00:40:01	Soft rock	27.9	40	04-Feb-77		9.5

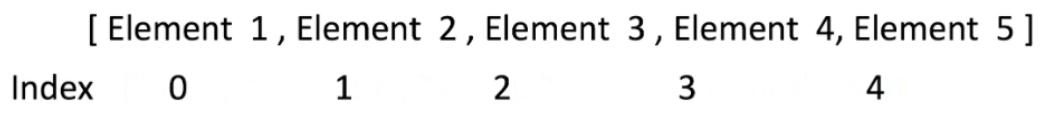
Lists

Indexing

We are going to take a look at lists in Python. A list is a sequenced collection of different objects such as integers, strings, and other lists as well. The address

of each element within a list is called an **index**. An index is used to access and refer to items within a list.



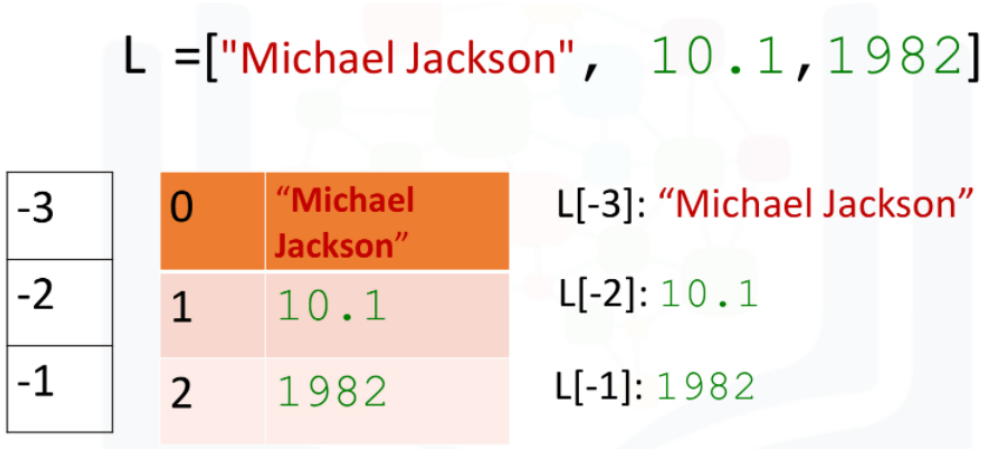


To create a list, type the list within square brackets **[ ]**, with your content inside the parenthesis and separated by commas. Let's try it!

In [1]:
```
# Create a list

L = ["Michael Jackson", 10.1, 1982]
L
```

Out[1]:
```
['Michael Jackson', 10.1, 1982]
```

We can use negative and regular indexing with a list :



In [2]:
```
# Print the elements on each index

print('the same element using negative and positive indexing:\n Postive:',L[0],
'\n Negative:' , L[-3]  )
print('the same element using negative and positive indexing:\n Postive:',L[1],
'\n Negative:' , L[-2]  )
print('the same element using negative and positive indexing:\n Postive:',L[2],
'\n Negative:' , L[-1]  )
```

```
the same element using negative and positive indexing:
 Postive: Michael Jackson
 Negative: Michael Jackson
the same element using negative and positive indexing:
 Postive: 10.1
 Negative: 10.1
the same element using negative and positive indexing:
 Postive: 1982
 Negative: 1982
```

List Content

Lists can contain strings, floats, and integers. We can nest other lists, and we can also nest tuples and other data structures. The same indexing conventions apply for nesting:

In [3]:
```
# Sample List
```

```
["Michael Jackson", 10.1, 1982, [1, 2], ("A", 1)]
```

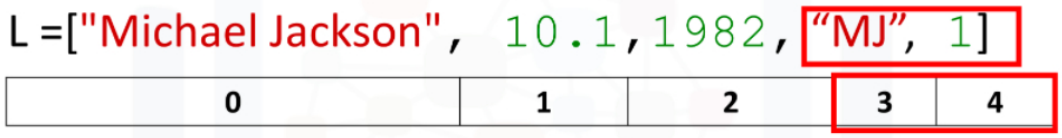
Out[3]:
```
['Michael Jackson', 10.1, 1982, [1, 2], ('A', 1)]
```

List Operations

We can also perform slicing in lists. For example, if we want the last two elements, we use the following command:

In [4]:
```
# Sample List

L = ["Michael Jackson", 10.1,1982,"MJ",1]
L
```

Out[4]:
```
['Michael Jackson', 10.1, 1982, 'MJ', 1]
```



In [5]:
```
# List slicing

L[3:5]
```

Out[5]:
```
['MJ', 1]
```

We can use the method `extend` to add new elements to the list:

In [6]:
```
# Use extend to add elements to list

L = [ "Michael Jackson", 10.2]
L.extend(['pop', 10])
L
```

Out[6]:
```
['Michael Jackson', 10.2, 'pop', 10]
```

Another similar method is `append`. If we apply `append` instead of `extend`, we add one element to the list:

In [7]:
```
# Use append to add elements to list

L = [ "Michael Jackson", 10.2]
L.append(['pop', 10])
L
```

Out[7]:
```
['Michael Jackson', 10.2, ['pop', 10]]
```

Each time we apply a method, the list changes. If we apply `extend` we add two new elements to the list. The list `L` is then modified by adding two new elements:

In [8]:
```
# Use extend to add elements to list

L = [ "Michael Jackson", 10.2]
L.extend(['pop', 10])
L
```

Out[8]:
```
['Michael Jackson', 10.2, 'pop', 10]
```

If we append the list `['a','b']` we have one new element consisting of a nested list:

In [9]:
```
# Use append to add elements to list

L.append(['a','b'])
L
```

Out[9]:
```
['Michael Jackson', 10.2, 'pop', 10, ['a', 'b']]
```

As lists are mutable, we can change them. For example, we can change the first element as follows:

In [10]:
```
# Change the element based on the index

A = ["disco", 10, 1.2]
print('Before change:', A)
A[0] = 'hard rock'
print('After change:', A)
```

```
Before change: ['disco', 10, 1.2]
After change: ['hard rock', 10, 1.2]
```

We can also delete an element of a list using the `del` command:

In [11]:
```
# Delete the element based on the index

print('Before change:', A)
del(A[0])
print('After change:', A)
```

```
Before change: ['hard rock', 10, 1.2]
After change: [10, 1.2]
```

We can convert a string to a list using `split`. For example, the method `split` translates every group of characters seperated by a space into an element in

We can convert a string to a list using `split`. For example, the method `split` translates every group of characters separated by a space into an element in a list:

In [12]:
```
# Split the string, default is by space

'hard rock'.split()
```

Out[12]:
```
['hard', 'rock']
```

We can use the split function to separate strings on a specific character. We pass the character we would like to split on into the argument, which in this case is a comma. The result is a list, and each element corresponds to a set of characters that have been separated by a comma:

In [13]:
```
# Split the string by comma

'A,B,C,D'.split(',')
```

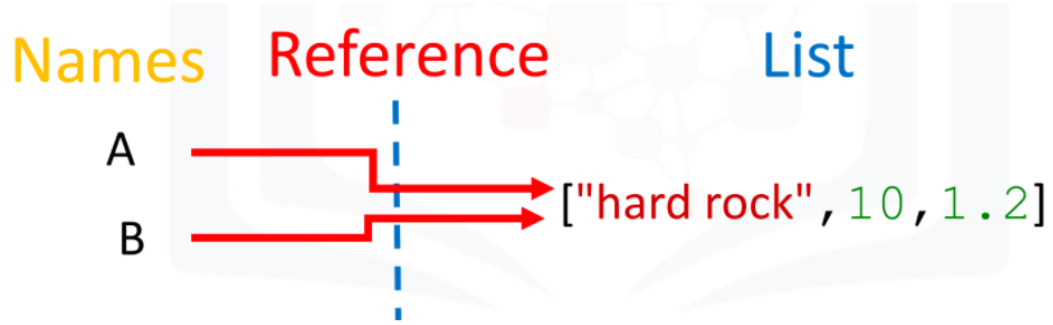
Out[13]:
```
['A', 'B', 'C', 'D']
```

Copy and Clone List

When we set one variable **B** equal to **A**; both **A** and **B** are referencing the same list in memory:

In [14]:
```
# Copy (copy by reference) the list A

A = ["hard rock", 10, 1.2]
B = A
print('A:', A)
print('B:', B)
```

```
A: ['hard rock', 10, 1.2]
B: ['hard rock', 10, 1.2]
```

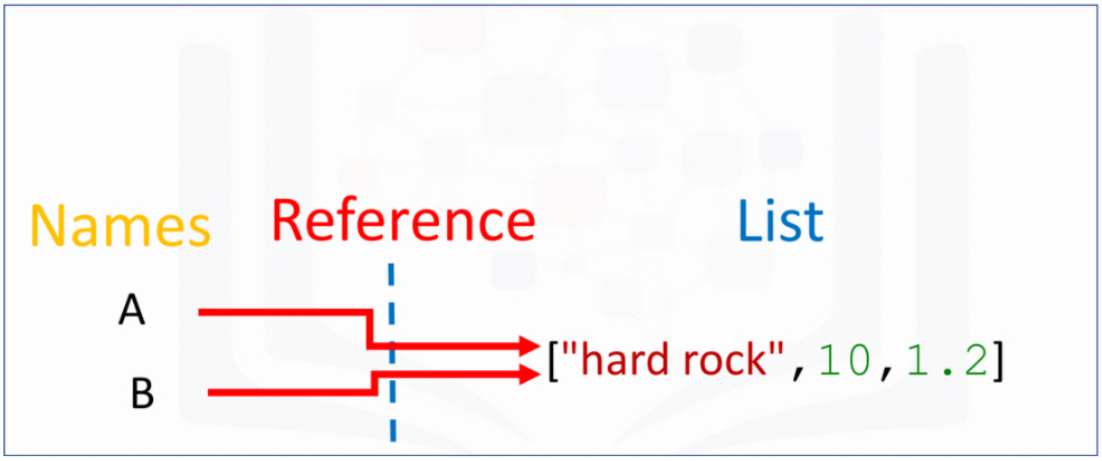


Initially, the value of the first element in **B** is set as hard rock. If we change the first element in **A** to **banana**, we get an unexpected side effect. As **A** and **B** are referencing the same list, if we change list **A**, then list **B** also changes. If we check the first element of **B** we get banana instead of hard rock:

In [15]:
```
# Examine the copy by reference

print('B[0]:', B[0])
A[0] = "banana"
print('B[0]:', B[0])
```

```
B[0]: hard rock
B[0]: banana
```

This is demonstrated in the following figure:



You can clone list **A** by using the following syntax:

In [16]:
```
# Clone (clone by value) the list A

B = A[:]
B
```

Out[16]:
```
['banana', 10, 1.2]
```

Variable **B** references a new copy or clone of the original list; this is demonstrated in the following figure:

Now if you change **A**, **B** will not change:

In [17]:
```
print('B[0]:', B[0])
A[0] = "hard rock"
print('B[0]:', B[0])
```

```
B[0]: banana
B[0]: banana
```

Quiz on List

Create a list `a_list`, with the following elements `1`, `hello`, `[1,2,3]` and `True`.

In [19]:
```
# Write your code below and press Shift+Enter to execute
a_list = [1, "hello", [1,2,3] ,"True"]
```

Double-click **here** for the solution.

Find the value stored at index 1 of `a_list`.

In [21]:
```
# Write your code below and press Shift+Enter to execute
a_list[1]
```

Out[21]:
```
'hello'
```

Double-click **here** for the solution.

Retrieve the elements stored at index 1, 2 and 3 of `a_list`.

In [23]:
```
# Write your code below and press Shift+Enter to execute
a_list[1:4]
```

Out[23]:
```
['hello', [1, 2, 3], 'True']
```

Double-click **here** for the solution.

A = [1, 'a']Concatenate the following lists `A = [1, 'a']` and `B = [2, 1, 'd']`:

In [24]:
```
# Write your code below and press Shift+Enter to execute
A = [1, 'a']
B = [2, 1, 'd']
A + B
```

Out[24]:
```
[1, 'a', 2, 1, 'd']
```

Double-click **here** for the solution.

The last exercise!

Congratulations, you have completed your first lesson and hands-on lab in Python. However, there is one more thing you need to do. The Data Science community encourages sharing work. The best way to share and showcase your work is to share it on GitHub. By sharing your notebook on GitHub you are not only building your reputation with fellow data scientists, but you can also show it off when applying for a job. Even though this was your first piece of work, it is never too early to start building good habits. So, please read and follow this article to learn how to share your work.

Author

Joseph Santarcangelo

Other contributors

Mavis Zhou

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2020-08-26	2.0	Lavanya	Moved lab to course repo in GitLab

Tuples in Python

Estaimted time needed: **15** minutes

Objectives

After compliting this lab you will be able to:

- Perform the basics tuple operations in Python, including indexing, slicing and sorting

Table of Contents

Estimated time needed: **15 min**

About the Dataset

Imagine you received album recommendations from your friends and compiled all of the recommendations into a table, with specific information about each album.

The table has one row for each movie and several columns:

- **artist** - Name of the artist
- **album** - Name of the album
- **released_year** - Year the album was released
- **length_min_sec** - Length of the album (hours,minutes,seconds)
- **genre** - Genre of the album
- **music_recording_sales_millions** - Music recording sales (millions in USD) on SONG://DATABASE
- **claimed_sales_millions** - Album's claimed sales (millions in USD) on SONG://DATABASE
- **date_released** - Date on which the album was released
- **soundtrack** - Indicates if the album is the movie soundtrack (Y) or (N)
- **rating_of_friends** - Indicates the rating from your friends from 1 to 10

The dataset can be seen below:

Artist	Album	Released	Length	Genre	Music recording sales (millions)	Claimed sales (millions)	Released	Soundtrack	Rating (friends)
Michael Jackson	Thriller	1982	00:42:19	Pop, rock, R&B	46	65	30-Nov-82		10.0
AC/DC	Back in Black	1980	00:42:11	Hard rock	26.1	50	25-Jul-80		8.5
Pink Floyd	The Dark Side of the Moon	1973	00:42:49	Progressive rock	24.2	45	01-Mar-73		9.5
Whitney Houston	The Bodyguard	1992	00:57:44	Soundtrack/R&B, soul, pop	26.1	50	25-Jul-80	Y	7.0
Meat Loaf	Bat Out of Hell	1977	00:46:33	Hard rock, progressive rock	20.6	43	21-Oct-77		7.0
Eagles	Their Greatest Hits (1971-1975)	1976	00:43:08	Rock, soft rock, folk rock	32.2	42	17-Feb-76		9.5
Bee Gees	Saturday Night Fever	1977	1:15:54	Disco	20.6	40	15-Nov-77	Y	9.0
Fleetwood Mac	Rumours	1977	00:40:01	Soft rock	27.9	40	04-Feb-77		9.5

Tuples

In Python, there are different data types: string, integer and float. These data types can all be contained in a tuple as follows:

Now, let us create your first tuple with string, integer and float.

In [1]:
```
# Create your first tuple

tuple1 = ("disco",10,1.2 )
tuple1
```

Out[1]:
```
('disco', 10, 1.2)
```

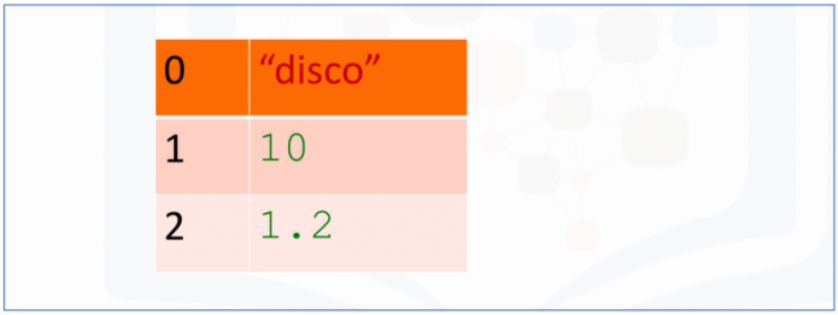
The type of variable is a **tuple**.

In [2]:
```
# Print the type of the tuple you created

type(tuple1)
```

Out[2]:
```
tuple
```

Indexing

Each element of a tuple can be accessed via an index. The following table represents the relationship between the index and the items in the tuple. Each element can be obtained by the name of the tuple followed by a square bracket with the index number:



We can print out each value in the tuple:

In [3]:
```
# Print the variable on each index

print(tuple1[0])
print(tuple1[1])
print(tuple1[2])
```

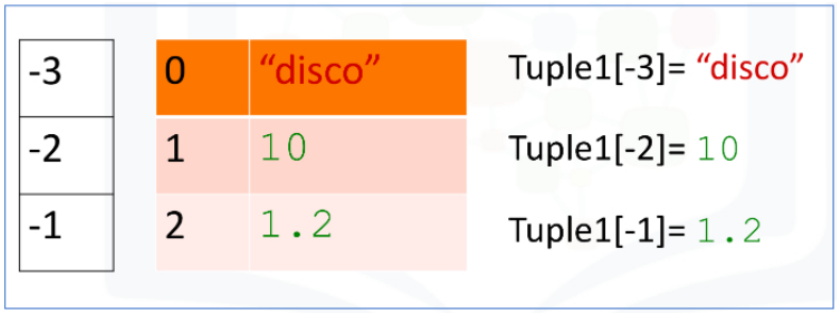
```
disco
10
1.2
```

We can print out the **type** of each value in the tuple:

In [4]:
```
# Print the type of value on each index

print(type(tuple1[0]))
print(type(tuple1[1]))
print(type(tuple1[2]))
```

```
<class 'str'>
<class 'int'>
<class 'float'>
```

We can also use negative indexing. We use the same table above with corresponding negative values:



We can obtain the last element as follows (this time we will not use the print statement to display the values):

In [5]:
```
# Use negative index to get the value of the last element

tuple1[-1]
```

Out[5]:
```
1.2
```

We can display the next two elements as follows:

In [6]:
```
# Use negative index to get the value of the second last element

tuple1[-2]
```

Out[6]:
```
10
```

In [7]:
```
# Use negative index to get the value of the third last element

tuple1[-3]
```

Out[7]:
```
'disco'
```

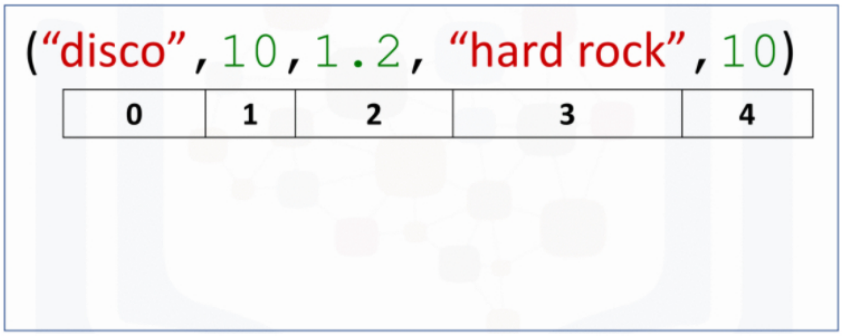
Concatenate Tuples

We can concatenate or combine tuples by using the **+** sign:

In [8]:
```
# Concatenate two tuples

tuple2 = tuple1 + ("hard rock", 10)
tuple2
```

Out[8]:
```
('disco', 10, 1.2, 'hard rock', 10)
```

We can slice tuples obtaining multiple values as demonstrated by the figure below:



Slicing

We can slice tuples, obtaining new tuples with the corresponding elements:

In [9]:
```
# Slice from index 0 to index 2

tuple2[0:3]
```

Out[9]:
```
('disco', 10, 1.2)
```

We can obtain the last two elements of the tuple:

In [10]:
```
# Slice from index 3 to index 4

tuple2[3:5]
```

Out[10]:
```
('hard rock', 10)
```

We can obtain the length of a tuple using the length command:

In [11]:
```
# Get the length of tuple

len(tuple2)
```

Out[11]:
```
5
```

This figure shows the number of elements:

Sorting

Consider the following tuple:

In [12]:
```
# A sample tuple

Ratings = (0, 9, 6, 5, 10, 8, 9, 6, 2)
```

We can sort the values in a tuple and save it to a new tuple:

In [13]:
```
# Sort the tuple

RatingsSorted = sorted(Ratings)
RatingsSorted
```

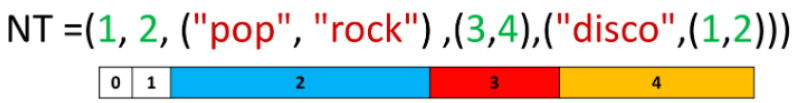
Out[13]:
```
[0, 2, 5, 6, 6, 8, 9, 9, 10]
```

Nested Tuple

A tuple can contain another tuple as well as other more complex data types. This process is called 'nesting'. Consider the following tuple with several elements:

In [14]:
```
# Create a nest tuple

NestedT =(1, 2, ("pop", "rock") ,(3,4),("disco",(1,2)))
```

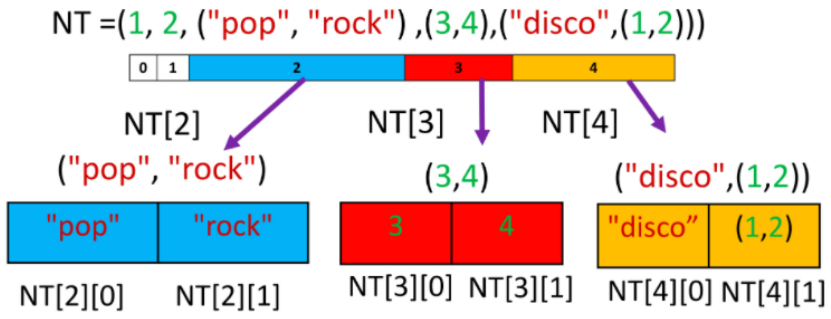
Each element in the tuple including other tuples can be obtained via an index as shown in the figure:



In [15]:
```
# Print element on each index

print("Element 0 of Tuple: ", NestedT[0])
print("Element 1 of Tuple: ", NestedT[1])
print("Element 2 of Tuple: ", NestedT[2])
print("Element 3 of Tuple: ", NestedT[3])
print("Element 4 of Tuple: ", NestedT[4])
```

```
Element 0 of Tuple:  1
Element 1 of Tuple:  2
Element 2 of Tuple:  ('pop', 'rock')
Element 3 of Tuple:  (3, 4)
Element 4 of Tuple:  ('disco', (1, 2))
```

We can use the second index to access other tuples as demonstrated in the figure:



We can access the nested tuples :

In [16]:
```
# Print element on each index, including nest indexes

print("Element 2, 0 of Tuple: ",   NestedT[2][0])
print("Element 2, 1 of Tuple: ",   NestedT[2][1])
print("Element 3, 0 of Tuple: ",   NestedT[3][0])
print("Element 3, 1 of Tuple: ",   NestedT[3][1])
print("Element 4, 0 of Tuple: ",   NestedT[4][0])
print("Element 4, 1 of Tuple: ",   NestedT[4][1])
```

```
Element 2, 0 of Tuple:  pop
Element 2, 1 of Tuple:  rock
Element 3, 0 of Tuple:  3
Element 3, 1 of Tuple:  4
Element 4, 0 of Tuple:  disco
Element 4, 1 of Tuple:  (1, 2)
```

We can access strings in the second nested tuples using a third index:

In [17]:
```
# Print the first element in the second nested tuples

NestedT[2][1][0]
```

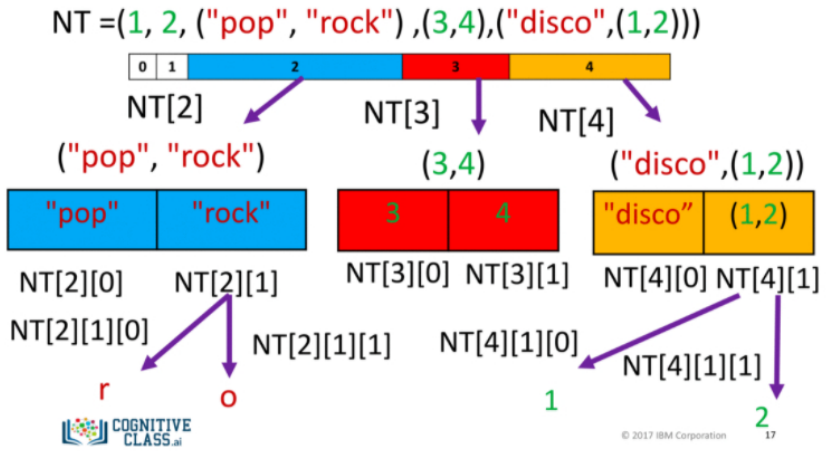
Out[17]:
```
'r'
```

In [18]:
```
# Print the second element in the second nested tuples
```

```
NestedT[2][1][1]
```

Out[18]: 'o'

We can use a tree to visualise the process. Each new index corresponds to a deeper level in the tree:



Similarly, we can access elements nested deeper in the tree with a fourth index:

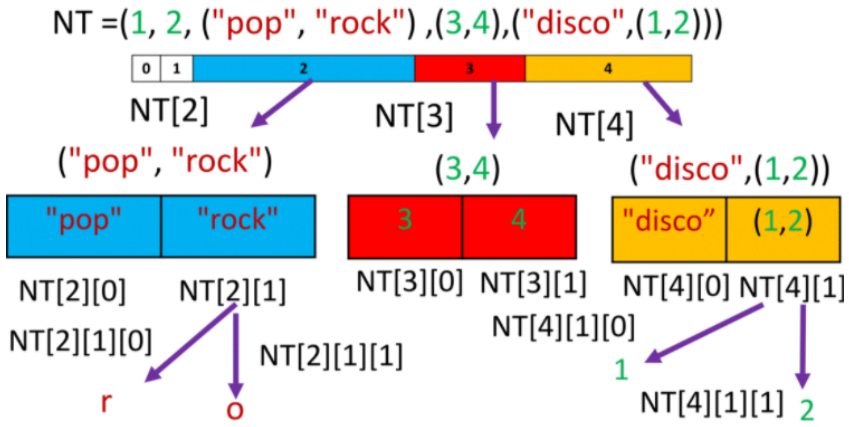
In [19]:
```
# Print the first element in the second nested tuples

NestedT[4][1][0]
```

Out[19]: 1

In [20]:
```
# Print the second element in the second nested tuples

NestedT[4][1][1]
```

Out[20]: 2

The following figure shows the relationship of the tree and the element `NestedT[4][1][1]`:



Quiz on Tuples

Consider the following tuple:

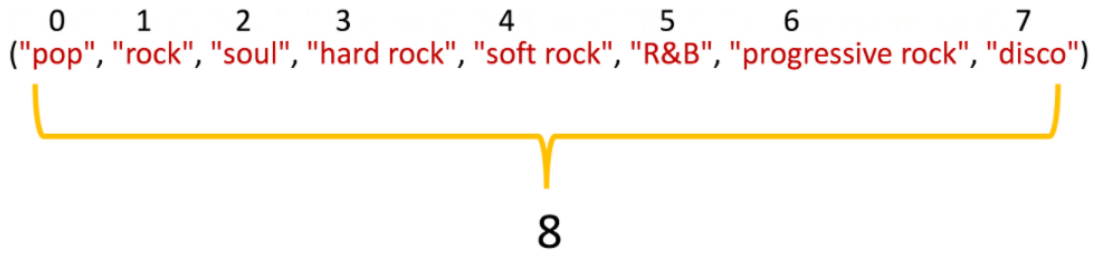
In [21]:
```
# sample tuple

genres_tuple = ("pop", "rock", "soul", "hard rock", "soft rock", \
                "R&B", "progressive rock", "disco")
genres_tuple
```

Out[21]:
```
('pop',
 'rock',
 'soul',
 'hard rock',
 'soft rock',
 'R&B',
 'progressive rock',
 'disco')
```

Find the length of the tuple, `genres_tuple`:

In [22]:
```
# Write your code below and press Shift+Enter to execute
len(genres_tuple)
```

Out[22]: 8



Double-click **here** for the solution.

Access the element, with respect to index 3:

In [24]:
```
# Write your code below and press Shift+Enter to execute
genres_tuple[3]
```

Out[24]:
```
'hard rock'
```

Double-click **here** for the solution.

Use slicing to obtain indexes 3, 4 and 5:

In [25]:
```
# Write your code below and press Shift+Enter to execute
genres_tuple[3:6]
```

Out[25]:
```
('hard rock', 'soft rock', 'R&B')
```

Double-click **here** for the solution.

Find the first two elements of the tuple `genres_tuple`:

In [28]:
```
# Write your code below and press Shift+Enter to execute
genres_tuple[0:2]
```

Out[28]:
```
('pop', 'rock')
```

Double-click **here** for the solution.

Find the first index of `"disco"`:

In [31]:
```
# Write your code below and press Shift+Enter to execute
genres_tuple.index("disco")
```

Out[31]:
```
7
```

Double-click **here** for the solution.

Generate a sorted List from the Tuple `C_tuple=(-5, 1, -3)`:

In [35]:
```
# Write your code below and press Shift+Enter to execute
C_tuple=(-5, 1, -3)
sorted_list = sorted(C_tuple)
sorted_list
```

Out[35]:
```
[-5, -3, 1]
```

Double-click **here** for the solution.

The last exercise!

Congratulations, you have completed your first lesson and hands-on lab in Python. However, there is one more thing you need to do. The Data Science community encourages sharing work. The best way to share and showcase your work is to share it on GitHub. By sharing your notebook on GitHub you are not only building your reputation with fellow data scientists, but you can also show it off when applying for a job. Even though this was your first piece of work, it is never too early to start building good habits. So, please read and follow this article to learn how to share your work.

Author

Joseph Santarcangelo

Other contributors

Mavis Zhou

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2020-08-26	2.0	Lavanya	Moved lab to course repo in GitLab

Dictionaries in Python

Estaimted time needed: **20** minutes

Objectives

After compliting this lab you will be able to:

- Work with libraries in Python, including operations

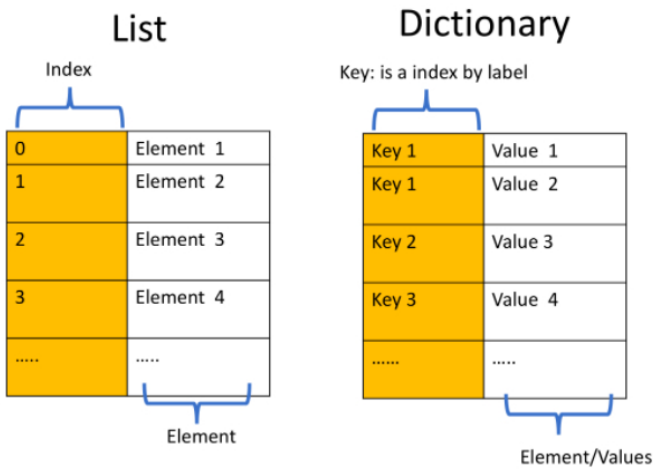
Table of Contents

Estimated time needed: **20 min**

Dictionaries

What are Dictionaries?

A dictionary consists of keys and values. It is helpful to compare a dictionary to a list. Instead of the numerical indexes such as a list, dictionaries have keys. These keys are the keys that are used to access values within a dictionary.



An example of a Dictionary `Dict`:

In [1]:
```
# Create the dictionary

Dict = {"key1": 1, "key2": "2", "key3": [3, 3, 3], "key4": (4, 4, 4), ('key5'): 5, (0, 1): 6}
Dict
```

Out[1]:
```
{'key1': 1,
 'key2': '2',
 'key3': [3, 3, 3],
 'key4': (4, 4, 4),
 'key5': 5,
 (0, 1): 6}
```

The keys can be strings:

In [2]:
```
# Access to the value by the key

Dict["key1"]
```

Out[2]:
```
1
```

Keys can also be any immutable object such as a tuple:

In [3]:
```
# Access to the value by the key

Dict[(0, 1)]
```

Out[3]: 6

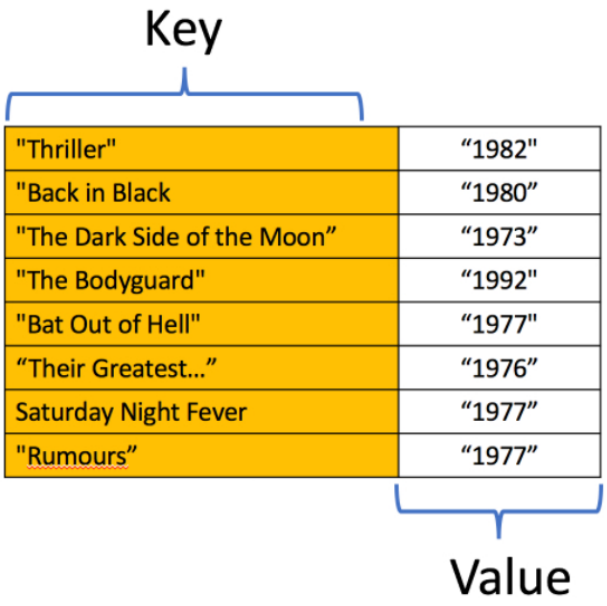
Each key is separated from its value by a colon " `:` ". Commas separate the items, and the whole dictionary is enclosed in curly braces. An empty dictionary without any items is written with just two curly braces, like this " `{}` ".

In [4]:
```
# Create a sample dictionary

release_year_dict = {"Thriller": "1982", "Back in Black": "1980", \
                     "The Dark Side of the Moon": "1973", "The Bodyguard": "1992", \
                     "Bat Out of Hell": "1977", "Their Greatest Hits (1971-1975)": "1976", \
                     "Saturday Night Fever": "1977", "Rumours": "1977"}
release_year_dict
```

Out[4]:
```
{'Thriller': '1982',
 'Back in Black': '1980',
 'The Dark Side of the Moon': '1973',
 'The Bodyguard': '1992',
 'Bat Out of Hell': '1977',
 'Their Greatest Hits (1971-1975)': '1976',
 'Saturday Night Fever': '1977',
 'Rumours': '1977'}
```

In summary, like a list, a dictionary holds a sequence of elements. Each element is represented by a key and its corresponding value. Dictionaries are created with two curly braces containing keys and values separated by a colon. For every key, there can only be one single value, however, multiple keys can hold the same value. Keys can only be strings, numbers, or tuples, but values can be any data type.

It is helpful to visualize the dictionary as a table, as in the following image. The first column represents the keys, the second column represents the values.



Keys

You can retrieve the values based on the names:

In [5]:
```
# Get value by keys

release_year_dict['Thriller']
```

Out[5]: '1982'

This corresponds to:

"Thriller"	"1982"
"Back in Black	"1980"
"The Dark Side of the Moon"	"1973"
"The Bodyguard"	"1992"
"Bat Out of Hell"	"1977"
"Their Greatest..."	"1976"
"Saturday Night Fever"	"1977"
"Rumours"	"1977"

Similarly for **The Bodyguard**

In [6]:
```
# Get value by key

release_year_dict['The Bodyguard']
```

Out[6]:
```
'1992'
```

"Thriller"	“1982"
"Back in Black	“1980”
"The Dark Side of the Moon”	“1973”
"The Bodyguard"	“1992"
"Bat Out of Hell"	“1977"
“Their Greatest...”	“1976”
“Saturday Night Fever”	“1977”
"Rumours”	“1977”

Now let you retrieve the keys of the dictionary using the method `release_year_dict()`:

In [7]:
```
# Get all the keys in dictionary

release_year_dict.keys()
```

Out[7]:
```
dict_keys(['Thriller', 'Back in Black', 'The Dark Side of the Moon', 'The Bodyguard', 'Bat Out of Hell', 'Their Greatest Hits
(1971-1975)', 'Saturday Night Fever', 'Rumours'])
```

You can retrieve the values using the method `values()`:

In [8]:
```
# Get all the values in dictionary

release_year_dict.values()
```

Out[8]:
```
dict_values(['1982', '1980', '1973', '1992', '1977', '1976', '1977', '1977'])
```

We can add an entry:

In [9]:
```
# Append value with key into dictionary

release_year_dict['Graduation'] = '2007'
release_year_dict
```

Out[9]:
```
{'Thriller': '1982',
 'Back in Black': '1980',
 'The Dark Side of the Moon': '1973',
 'The Bodyguard': '1992',
 'Bat Out of Hell': '1977',
 'Their Greatest Hits (1971-1975)': '1976',
 'Saturday Night Fever': '1977',
 'Rumours': '1977',
 'Graduation': '2007'}
```

We can delete an entry:

In [10]:
```
# Delete entries by key

del(release_year_dict['Thriller'])
del(release_year_dict['Graduation'])
release_year_dict
```

Out[10]:
```
{'Back in Black': '1980',
 'The Dark Side of the Moon': '1973',
 'The Bodyguard': '1992',
 'Bat Out of Hell': '1977',
 'Their Greatest Hits (1971-1975)': '1976',
 'Saturday Night Fever': '1977',
 'Rumours': '1977'}
```

We can verify if an element is in the dictionary:

In [11]:
```
# Verify the key is in the dictionary

'The Bodyguard' in release_year_dict
```

Out[11]:
```
True
```

Quiz on Dictionaries

You will need this dictionary for the next two questions:

In [12]:
```
# Question sample dictionary
```

```
soundtrack_dic = {"The Bodyguard":"1992", "Saturday Night Fever":"1977"}
soundtrack_dic
```

Out[12]:
```
{'The Bodyguard': '1992', 'Saturday Night Fever': '1977'}
```

a) In the dictionary `soundtrack_dic` what are the keys ?

In [14]:
```
soundtrack_dic.keys()
```

Out[14]:
```
dict_keys(['The Bodyguard', 'Saturday Night Fever'])
```

Double-click **here** for the solution.

b) In the dictionary `soundtrack_dic` what are the values ?

In [15]:
```
soundtrack_dic.values()
```

Out[15]:
```
dict_values(['1992', '1977'])
```

Double-click **here** for the solution.

You will need this dictionary for the following questions:

The Albums **Back in Black**, **The Bodyguard** and **Thriller** have the following music recording sales in millions 50, 50 and 65 respectively:

a) Create a dictionary `album_sales_dict` where the keys are the album name and the sales in millions are the values.

In [17]:
```
album_sales_dict = {"Back in Black":50, "The Bodyguard": 50, "Thriller":65}
```

Double-click **here** for the solution.

b) Use the dictionary to find the total sales of **Thriller**:

In [18]:
```
album_sales_dict["Thriller"]
```

Out[18]:
```
65
```

Double-click **here** for the solution.

c) Find the names of the albums from the dictionary using the method `keys` :

In [19]:
```
album_sales_dict.keys()
```

Out[19]:
```
dict_keys(['Back in Black', 'The Bodyguard', 'Thriller'])
```

Double-click **here** for the solution.

d) Find the names of the recording sales from the dictionary using the method `values` :

In [21]:
```
album_sales_dict.values()
```

Out[21]:
```
dict_values([50, 50, 65])
```

Double-click **here** for the solution.

The last exercise!

Congratulations, you have completed your first lesson and hands-on lab in Python. However, there is one more thing you need to do. The Data Science community encourages sharing work. The best way to share and showcase your work is to share it on GitHub. By sharing your notebook on GitHub you are not only building your reputation with fellow data scientists, but you can also show it off when applying for a job. Even though this was your first piece of work, it is never too early to start building good habits. So, please read and follow this article to learn how to share your work.

Author

Joseph Santarcangelo

Other contributors

Mavis Zhou

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description

2020-09-09	2.1	Malika Singla	Updated the variable soundtrack_dict to soundtrack_dic in Questions
2020-08-26	2.0	Lavanya	Moved lab to course repo in GitLab

In [ ]:

Sets in Python

Estaimted time needed: **20** minutes

Objectives

After compliting this lab you will be able to:

- Work with sets in Python, including operations and logic operations.

Table of Contents

Estimated time needed: **20 min**

Sets

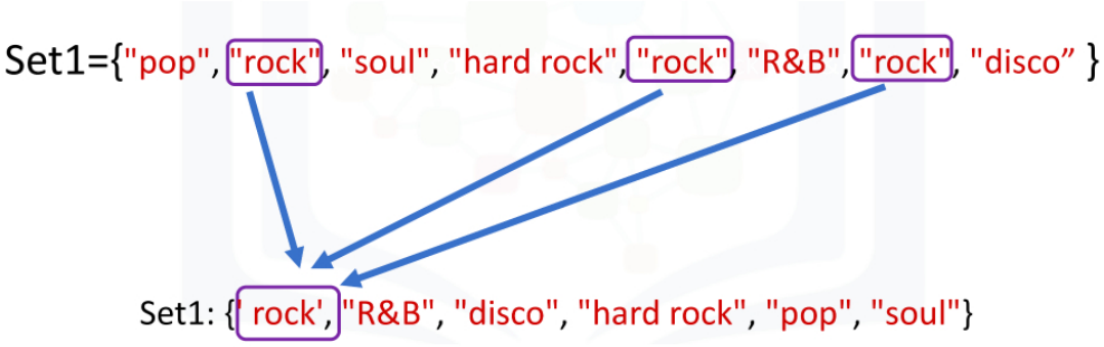
Set Content

A set is a unique collection of objects in Python. You can denote a set with a curly bracket **{}**. Python will automatically remove duplicate items:

In [1]:
```
# Create a set

set1 = {"pop", "rock", "soul", "hard rock", "rock", "R&B", "rock", "disco"}
set1
```

Out[1]:
```
{'R&B', 'disco', 'hard rock', 'pop', 'rock', 'soul'}
```

The process of mapping is illustrated in the figure:



You can also create a set from a list as follows:

In [2]:
```
# Convert list to set

album_list = [ "Michael Jackson", "Thriller", 1982, "00:42:19", \
               "Pop, Rock, R&B", 46.0, 65, "30-Nov-82", None, 10.0]
album_set = set(album_list)
album_set
```

Out[2]:
```
{'00:42:19',
 10.0,
 1982,
 '30-Nov-82',
 46.0,
 65,
 'Michael Jackson',
 None,
 'Pop, Rock, R&B',
 'Thriller'}
```

Now let us create a set of genres:

In [3]:
```
# Convert list to set

music_genres = set(["pop", "pop", "rock", "folk rock", "hard rock", "soul", \
                    "progressive rock", "soft rock", "R&B", "disco"])
music_genres
```

Out[3]:
```
{'R&B',
 'disco',
 'folk rock',
 'hard rock',
 'pop',
 'progressive rock',
 'rock',
 'soft rock',
 'soul'}
```

Set Operations

Let us go over set operations, as these can be used to change the set. Consider the set **A**:

In [4]:
```
# Sample set

A = set(["Thriller", "Back in Black", "AC/DC"])
A
```

Out[4]:
```
{'AC/DC', 'Back in Black', 'Thriller'}
```

We can add an element to a set using the `add()` method:

In [5]:
```
# Add element to set

A.add("NSYNC")
A
```

Out[5]:
```
{'AC/DC', 'Back in Black', 'NSYNC', 'Thriller'}
```

If we add the same element twice, nothing will happen as there can be no duplicates in a set:

In [6]:
```
# Try to add duplicate element to the set

A.add("NSYNC")
A
```

Out[6]:
```
{'AC/DC', 'Back in Black', 'NSYNC', 'Thriller'}
```

We can remove an item from a set using the `remove` method:

In [7]:
```
# Remove the element from set

A.remove("NSYNC")
A
```

Out[7]:
```
{'AC/DC', 'Back in Black', 'Thriller'}
```

We can verify if an element is in the set using the `in` command:

In [8]:
```
# Verify if the element is in the set

"AC/DC" in A
```

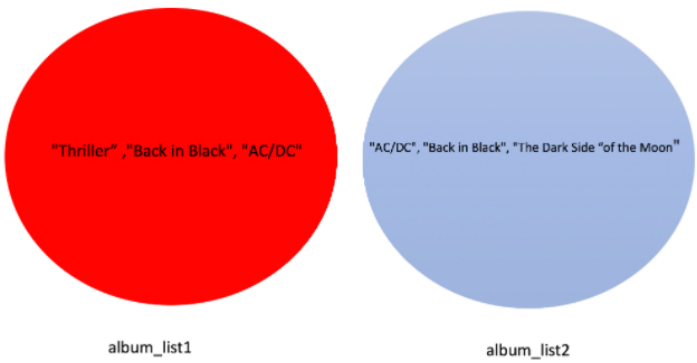
Out[8]:
```
True
```

Sets Logic Operations

Remember that with sets you can check the difference between sets, as well as the symmetric difference, intersection, and union:

Consider the following two sets:

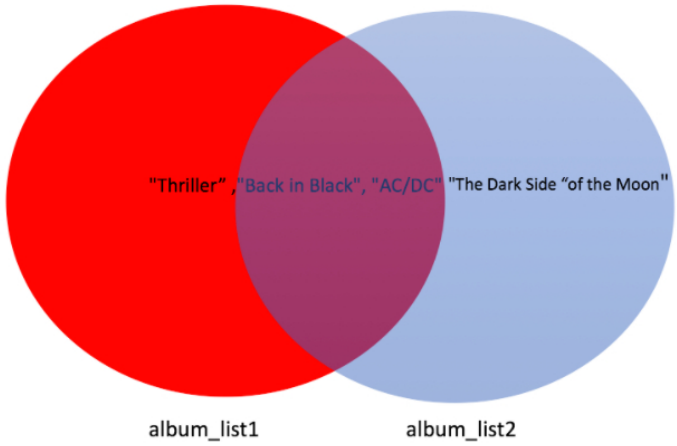
In [9]:
```
# Sample Sets

album_set1 = set(["Thriller", 'AC/DC', 'Back in Black'])
album_set2 = set([ "AC/DC", "Back in Black", "The Dark Side of the Moon"])
```



In [10]:
```
# Print two sets

album_set1, album_set2
```

Out[10]:
```
({'AC/DC', 'Back in Black', 'Thriller'},
 {'AC/DC', 'Back in Black', 'The Dark Side of the Moon'})
```

As both sets contain **AC/DC** and **Back in Black** we represent these common elements with the intersection of two circles.



You can find the intersect of two sets as follow using `&`:

In [11]:
```
# Find the intersections

intersection = album_set1 & album_set2
intersection
```

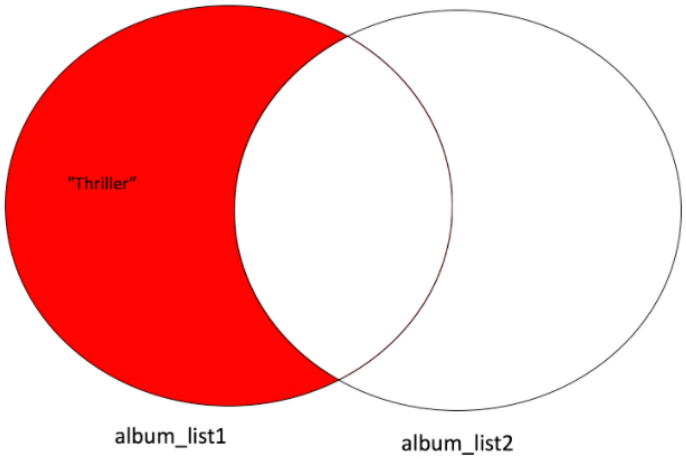
Out[11]:
```
{'AC/DC', 'Back in Black'}
```

You can find all the elements that are only contained in `album_set1` using the `difference` method:

In [12]:
```
# Find the difference in set1 but not set2

album_set1.difference(album_set2)
```

Out[12]:
```
{'Thriller'}
```

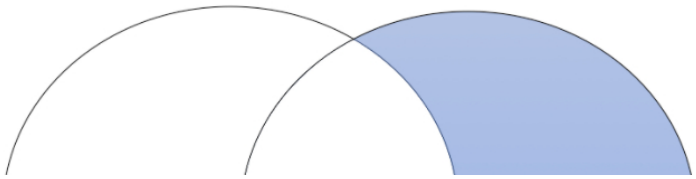
You only need to consider elements in `album_set1`; all the elements in `album_set2`, including the intersection, are not included.

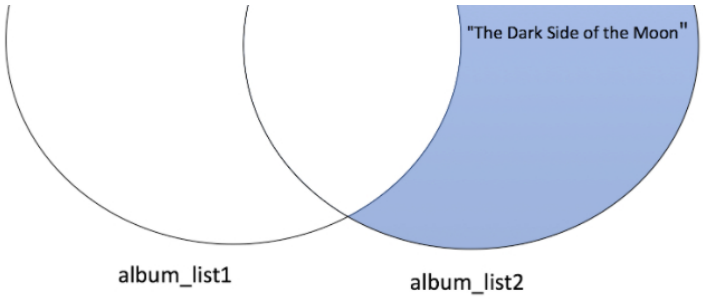


The elements in `album_set2` but not in `album_set1` is given by:

In [13]:
```
album_set2.difference(album_set1)
```

Out[13]:
```
{'The Dark Side of the Moon'}
```



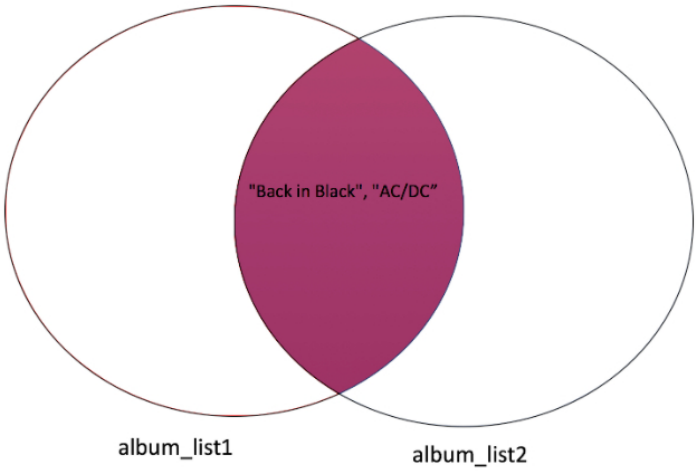


You can also find the intersection of `album_list1` and `album_list2`, using the `intersection` method:

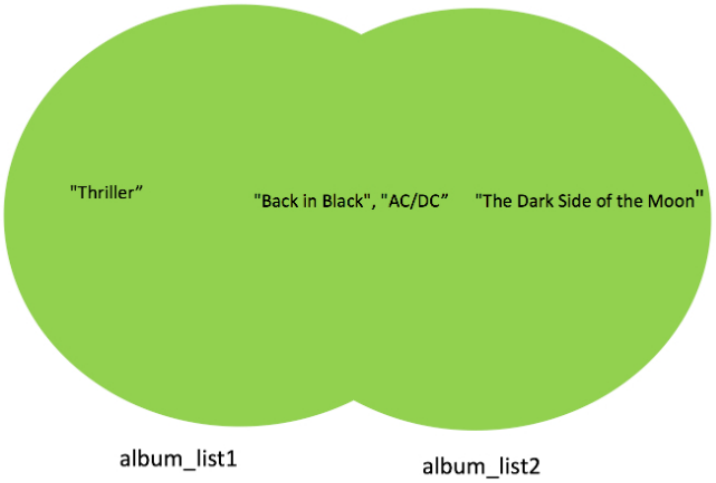
In [14]:
```
# Use intersection method to find the intersection of album_list1 and album_list2

album_set1.intersection(album_set2)
```

Out[14]:
```
{'AC/DC', 'Back in Black'}
```

This corresponds to the intersection of the two circles:



The union corresponds to all the elements in both sets, which is represented by coloring both circles:



The union is given by:

In [15]:
```
# Find the union of two sets

album_set1.union(album_set2)
```

Out[15]:
```
{'AC/DC', 'Back in Black', 'The Dark Side of the Moon', 'Thriller'}
```

And you can check if a set is a superset or subset of another set, respectively, like this:

In [16]:
```
# Check if superset

set(album_set1).issuperset(album_set2)
```

Out[16]: False

In [17]:
```
# Check if subset

set(album_set2).issubset(album_set1)
```

Out[17]: False

Here is an example where `issubset()` and `issuperset()` return true:

In [18]:
```
# Check if subset

set({"Back in Black", "AC/DC"}).issubset(album_set1)
```

Out[18]: True

In [19]:
```
# Check if superset

album_set1.issuperset({"Back in Black", "AC/DC"})
```

Out[19]: True

Quiz on Sets

Convert the list `['rap','house','electronic music', 'rap']` to a set:

In [21]:
```
# Write your code below and press Shift+Enter to execute
list = ['rap','house','electronic music', 'rap']
set_list = set(list)
```

Double-click **here** for the solution.

Consider the list `A = [1, 2, 2, 1]` and set `B = set([1, 2, 2, 1])`, does `sum(A) = sum(B)`

In [23]:
```
# Write your code below and press Shift+Enter to execute
A = [1, 2, 2, 1]
B = set([1, 2, 2, 1])
print("the sum of A is:", sum(A))
print("the sum of B is:", sum(B))
```

```
the sum of A is: 6
the sum of B is: 3
```

Double-click **here** for the solution.

Create a new set `album_set3` that is the union of `album_set1` and `album_set2`:

In [25]:
```
# Write your code below and press Shift+Enter to execute

album_set1 = set(["Thriller", 'AC/DC', 'Back in Black'])
album_set2 = set([ "AC/DC", "Back in Black", "The Dark Side of the Moon"])
album_set3 = album_set1.union(album_set2)
```

Double-click **here** for the solution.

Find out if `album_set1` is a subset of `album_set3`:

In [26]:
```
# Write your code below and press Shift+Enter to execute
album_set1.issubset(album_set3)
```

Out[26]: True

Double-click **here** for the solution.

The last exercise!

Congratulations, you have completed your first lesson and hands-on lab in Python. However, there is one more thing you need to do. The Data Science community encourages sharing work. The best way to share and showcase your work is to share it on GitHub. By sharing your notebook on GitHub you are not only building your reputation with fellow data scientists, but you can also show it off when applying for a job. Even though this was your first piece of work, it is never too early to start building good habits. So, please read and follow this article to learn how to share your work.

Author

Joseph Santarcangelo

Other contributors

Mavis Zhou

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2020-08-26	2.0	Lavanya	Moved lab to course repo in GitLab

In [ ]: