

# Python 3 - Strings

Strings are amongst the most popular types in Python. We can create them simply by enclosing characters in quotes. Python treats single quotes the same as double quotes. Creating strings is as simple as assigning a value to a variable. For example –

```
var1 = 'Hello World!'
var2 = "Python Programming"
```

## Accessing Values in Strings

Python does not support a character type; these are treated as strings of length one, thus also considered a substring.

To access substrings, use the square brackets for slicing along with the index or indices to obtain your substring. For example –

```
#!/usr/bin/python3

var1 = 'Hello World!'
var2 = "Python Programming"

print ("var1[0]: ", var1[0])
print ("var2[1:5]: ", var2[1:5])
```

[Live Demo](#)

When the above code is executed, it produces the following result –

```
var1[0]: H
var2[1:5]: ytho
```

## Updating Strings

You can "update" an existing string by (re)assigning a variable to another string. The new value can be related to its previous value or to a completely different string altogether. For example –

```
#!/usr/bin/python3
```

[Live Demo](#)

```
var1 = 'Hello World!'
print ("Updated String :- ", var1[:6] + 'Python')
```

When the above code is executed, it produces the following result –

```
Updated String :- Hello Python
```

## Escape Characters

Following table is a list of escape or non-printable characters that can be represented with backslash notation.

An escape character gets interpreted; in a single quoted as well as double quoted strings.

Backslash notation	Hexadecimal character	Description
\a	0x07	Bell or alert
\b	0x08	Backspace
\cx		Control-x
\C-x		Control-x
\e	0x1b	Escape
\f	0x0c	Formfeed
\M-\C-x		Meta-Control-x
\n	0x0a	Newline
\nnn		Octal notation, where n is in the range 0-7
\r	0x0d	Carriage return
\s	0x20	Space
\t	0x09	Tab
\v	0x0b	Vertical tab
\x		Character x
\xnn		Hexadecimal notation, where n is in the range 0-9, a-f, or A-F

## String Special Operators

Assume string variable **a** holds 'Hello' and variable **b** holds 'Python', then –

Operator	Description	Example
+	Concatenation - Adds values on either side of the operator	a + b will give HelloPython
*	Repetition - Creates new strings, concatenating multiple copies of the same string	a*2 will give -HelloHello
[]	Slice - Gives the character from the given index	a[1] will give e
[ : ]	Range Slice - Gives the characters from the given range	a[1:4] will give ell
in	Membership - Returns true if a character exists in the given string	H in a will give 1
not in	Membership - Returns true if a character does not exist in the given string	M not in a will give 1
r/R	Raw String - Suppresses actual meaning of Escape characters. The syntax for raw strings is exactly the same as for normal strings with the exception of the raw string operator, the letter "r," which precedes the quotation marks. The "r" can be lowercase (r) or uppercase (R) and must be placed immediately preceding the first quote mark.	print r'\n' prints \n and print R'\n'prints \n
%	Format - Performs String formatting	See at next section

## String Formatting Operator

One of Python's coolest features is the string format operator %. This operator is unique to strings and makes up for the pack of having functions from C's printf() family. Following is a simple example –

[Live Demo](#)

```
#!/usr/bin/python3

print ("My name is %s and weight is %d kg!" % ('Zara', 21))
```

When the above code is executed, it produces the following result –

```
My name is Zara and weight is 21 kg!
```

Here is the list of complete set of symbols which can be used along with % –

Sr.No.	Format Symbol & Conversion
1	<b>%c</b> character
2	<b>%s</b> string conversion via str() prior to formatting
3	<b>%i</b> signed decimal integer
4	<b>%d</b> signed decimal integer
5	<b>%u</b> unsigned decimal integer
6	<b>%o</b> octal integer
7	<b>%x</b> hexadecimal integer (lowercase letters)
8	<b>%X</b> hexadecimal integer (UPPERcase letters)
9	<b>%e</b> exponential notation (with lowercase 'e')
10	<b>%E</b> exponential notation (with UPPERcase 'E')

Other supported symbols and functionality are listed in the following table –

Sr.No.	Symbol & Functionality
1	<b>*</b> argument specifies width or precision
2	<b>-</b> left justification
3	<b>+</b> display the sign
4	<b>&lt;sp&gt;</b> leave a blank space before a positive number
5	<b>#</b> add the octal leading zero ( '0' ) or hexadecimal leading '0x' or '0X', depending on whether 'x' or 'X' were used.
6	<b>0</b> pad from left with zeros (instead of spaces)
7	<b>%</b> '%%%' leaves you with a single literal '%'
8	<b>(var)</b> mapping variable (dictionary arguments)
9	<b>m.n.</b> m is the minimum total width and n is the number of digits to display after the decimal point (if appl.)

## Triple Quotes

Python's triple quotes comes to the rescue by allowing strings to span multiple lines, including verbatim NEWLINES, TABs, and any other special characters.

The syntax for triple quotes consists of three consecutive **single or double** quotes.

```
#!/usr/bin/python3
```

Live Demo

```
para_str = """this is a long string that is made up of
several lines and non-printable characters such as
TAB ( \t ) and they will show up that way when displayed.
NEWLINES within the string, whether explicitly given like
this within the brackets [ \n ], or just a NEWLINE within
the variable assignment will also show up.
"""
print (para_str)
```

When the above code is executed, it produces the following result. Note how every single special character has been converted to its printed form, right down to the last NEWLINE at the end of the string between the "up." and closing triple quotes. Also note that NEWLINES occur either with an explicit carriage return at the end of a line or its escape code (\n) –

```
this is a long string that is made up of
several lines and non-printable characters such as
TAB (   ) and they will show up that way when displayed.
NEWLINES within the string, whether explicitly given like
this within the brackets [
 ], or just a NEWLINE within
the variable assignment will also show up.
```

Raw strings do not treat the backslash as a special character at all. Every character you put into a raw string stays the way you wrote it –

```
#!/usr/bin/python3
```

Live Demo

```
print ('C:\\nowhere')
```

When the above code is executed, it produces the following result –

```
C:\nowhere
```

Now let's make use of raw string. We would put expression in **r'expression'** as follows –

```
#!/usr/bin/python3  
  
print (r'C:\\nowhere')
```

Live Demo

When the above code is executed, it produces the following result –

```
C:\\nowhere
```

## Unicode String

In Python 3, all strings are represented in Unicode. In Python 2 are stored internally as 8-bit ASCII, hence it is required to attach 'u' to make it Unicode. It is no longer necessary now.

## Built-in String Methods

Python includes the following built-in methods to manipulate strings –

Sr.No.	Methods & Description
1	<code>capitalize()</code> Capitalizes first letter of string
2	<code>center(width, fillchar)</code> Returns a string padded with <i>fillchar</i> with the original string centered to a total of <i>width</i> columns.
3	<code>count(str, beg = 0, end = len(string))</code> Counts how many times <i>str</i> occurs in <i>string</i> or in a substring of <i>string</i> if starting index <i>beg</i> and ending index <i>end</i> are given.
4	<code>decode(encoding = 'UTF-8', errors = 'strict')</code> Decodes the string using the codec registered for encoding. <i>encoding</i> defaults to the default string encoding.
5	<code>encode(encoding = 'UTF-8', errors = 'strict')</code> Returns encoded string version of <i>string</i> ; on error, default is to raise a <code>ValueError</code> unless <i>errors</i> is given with 'ignore' or 'replace'.
6	<code>endswith(suffix, beg = 0, end = len(string))</code> Determines if <i>string</i> or a substring of <i>string</i> (if starting index <i>beg</i> and ending index <i>end</i> are given) ends with <i>suffix</i> ; returns true if so and false otherwise.
7	<code>expandtabs(tabsize = 8)</code> Expands tabs in <i>string</i> to multiple spaces; defaults to 8 spaces per tab if <i>tabsize</i> not provided.
8	<code>find(str, beg = 0, end = len(string))</code> Determine if <i>str</i> occurs in <i>string</i> or in a substring of <i>string</i> if starting index <i>beg</i> and ending index <i>end</i> are given returns index if found and -1 otherwise.
9	<code>index(str, beg = 0, end = len(string))</code> Same as <code>find()</code> , but raises an exception if <i>str</i> not found.

10	<code>isalnum()</code> Returns true if string has at least 1 character and all characters are alphanumeric and false otherwise.
11	<code>isalpha()</code> Returns true if string has at least 1 character and all characters are alphabetic and false otherwise.
12	<code>isdigit()</code> Returns true if string contains only digits and false otherwise.
13	<code>islower()</code> Returns true if string has at least 1 cased character and all cased characters are in lowercase and false otherwise.
14	<code>isnumeric()</code> Returns true if a unicode string contains only numeric characters and false otherwise.
15	<code>isspace()</code> Returns true if string contains only whitespace characters and false otherwise.
16	<code>istitle()</code> Returns true if string is properly "titlecased" and false otherwise.
17	<code>isupper()</code> Returns true if string has at least one cased character and all cased characters are in uppercase and false otherwise.
18	<code>join(seq)</code> Merges (concatenates) the string representations of elements in sequence seq into a string, with separator string.
19	<code>len(string)</code> Returns the length of the string

20	<code>ljust(width[, fillchar])</code> Returns a space-padded string with the original string left-justified to a total of width columns.
21	<code>lower()</code> Converts all uppercase letters in string to lowercase.
22	<code>lstrip()</code> Removes all leading whitespace in string.
23	<code>maketrans()</code> Returns a translation table to be used in translate function.
24	<code>max(str)</code> Returns the max alphabetical character from the string str.
25	<code>min(str)</code> Returns the min alphabetical character from the string str.
26	<code>replace(old, new [, max])</code> Replaces all occurrences of old in string with new or at most max occurrences if max given.
27	<code>rfind(str, beg = 0, end = len(string))</code> Same as find(), but search backwards in string.
28	<code>rindex( str, beg = 0, end = len(string))</code> Same as index(), but search backwards in string.
29	<code>rjust(width[, fillchar])</code> Returns a space-padded string with the original string right-justified to a total of width columns.
30	<code>rstrip()</code> Removes all trailing whitespace of string.