

# Python R training course - Session 2

Van-Duyet LE, Thinh PHAM

August 17, 2017

## Lecture list

1. Introduction to Python
2. **The Python Standard Library, if/loops statements**
  - Some useful and basic built-in functions (abs, etc.)
  - How to use import to import libraries (ex: math, random). We will introduce you several basic libraries that are often used in Python.
  - Introduce Comparison and Boolean Operators (if/else/while/for)
  - Students are also introduced the concept of object-orientation in Python.
3. Vector/matrix structures, numpy library
4. Python data types, File Processing, Pandas library
5. Functions in Python, Debugging.
6. Introduction to R, R for Python programmers.
7. Import data, plot data.
8. Data Mining in Python/R.

## 1 Some useful and basic built-in functions

List build-in functions: <https://docs.python.org/2/library/functions.html>

- **str()** is very common and useful for converting to strings, e.g. the number 1 to "1" -- similarly **int()** for converting other numbers (float or decimal), or strings, to integers.
- **len()** is perhaps most common, we often need to know how many members there are e.g. in a list and **len(mylist)** is the Python way for that (not like **mylist.length** in Javascript)
- **open()** is essential as that's what you use to open a file.
- **min()** and **max()** are common. also **all()** and **any()** are handy to know for basic logic.
- **set()** is useful because the set type doesn't have a literal like **[]** and **{}** for lists and dicts. **list()** is useful and quite common for converting e.g. a set to a list.
- **range()** for iterating number of times: **for i in range(10)**
- **sorted()** - creates sorted copy of list
- **help([object])** - To invoke the built-in help.

Built-in Functions				
abs()	divmod()	input()	open()	staticmethod()
all()	enumerate()	int()	ord()	str()
any()	eval()	isinstance()	pow()	sum()
basestring()	execfile()	issubclass()	print()	super()
bin()	file()	iter()	property()	tuple()
bool()	filter()	len()	range()	type()
bytearray()	float()	list()	raw_input()	unichr()
callable()	format()	locals()	reduce()	unicode()
chr()	frozenset()	long()	reload()	vars()
classmethod()	getattr()	map()	repr()	xrange()
cmp()	globals()	max()	reversed()	zip()
compile()	hasattr()	memoryview()	round()	__import__()
complex()	hash()	min()	set()	
delattr()	help()	next()	setattr()	
dict()	hex()	object()	slice()	
dir()	id()	oct()	sorted()	

## 1.1 Help

```
In [1]: help(range)
```

Help on built-in function range in module `__builtin__`:

```
range(...)
    range(stop) -> list of integers
    range(start, stop[, step]) -> list of integers
```

Return a list containing an arithmetic progression of integers.  
`range(i, j)` returns `[i, i+1, i+2, ..., j-1]`; start (!) defaults to 0.  
 When step is given, it specifies the increment (or decrement).  
 For example, `range(4)` returns `[0, 1, 2, 3]`. The end point is omitted!  
 These are exactly the valid indices for a list of 4 elements.

## 1.2 Type conversion & Type coercion

Python provides a collection of built-in functions that convert values from one type to another.

```
>>> int(5.6)
5

>>> float(6)
6.0

>>> 'Two, ' + str(2)
'Two, 2'

>>> bool(6)
True
...
```

### 1.3 Math

```
In [2]: max(91, 100)
```

```
Out[2]: 100
```

```
In [3]: min(57, min(24, 44))
```

```
Out[3]: 24
```

```
In [4]: for i in range(5, 10):  
        print i ** 2
```

```
25
```

```
36
```

```
49
```

```
64
```

```
81
```

```
In [5]: range(5,20, 2)
```

```
Out[5]: [5, 7, 9, 11, 13, 15, 17, 19]
```

```
In [6]: x = [1, 2, 3, 4]  
        y = ['a', 'b', 'c']
```

```
        zip(x, y)
```

```
Out[6]: [(1, 'a'), (2, 'b'), (3, 'c')]
```

```
In [6]: sorted([5,3,2,5,7,8])
```

```
Out[6]: [2, 3, 5, 5, 7, 8]
```

### 1.4 Strings

We have seen three types: int, float, and string. **Strings** are qualitatively different from the other two because they are made up of smaller pieces - characters.

The bracket operator selects a single character from a string.

```
In [7]: fruit = "banana"  
        letter = fruit[1]  
        print letter
```

```
a
```

```
In [7]: len([1,2])
```

```
Out[7]: 2
```

```
In [9]: for i in range(len(fruit)):
        print 'Char at', i, 'is', fruit[i]
```

```
Char at 0 is b
Char at 1 is a
Char at 2 is n
Char at 3 is a
Char at 4 is n
Char at 5 is a
```

### String slices

The "slice" syntax is a handy way to refer to sub-parts of sequences -- typically strings and lists. The slice `s[start:end]` is the elements beginning at `start` and extending up to but not including `end`. Suppose we have `s = "Hello"`

H	e	l	l	o
0	1	2	3	4
-5	-4	-3	-2	-1

```
In [9]: s = "Hello"
```

```
In [18]: print s[-1:-4]
```

```
oll
```

```
In [27]: s[4:1:-1]
```

```
Out[27]: 'oll'
```

```
In [28]: s[-4:-1]
```

```
Out[28]: 'ell'
```

```
In [25]: s[::-1]
```

```
Out[25]: 'olle'
```

```
In [29]: print s[-1]
```

11

```
In [5]: print s[-4:]
```

ello

```
In [7]: s[::-1]
```

```
Out[7]: 'olleH'
```

### Exercise

Ex1. Write a Python program to add 'ing' at the end of a given string (length should be at least 3). If the given string already ends with 'ing' then add 'ly' instead. If the string length of the given string is less than 3, leave it unchanged.

```
def add_string(s):  
    # your code goes here
```

```
>>> add_string('abc')  
    'abcing'
```

```
>>> add_string('string')  
    'stringly'
```

```
In [2]: # Solution  
        def add_string(s):  
            if len(s) < 3: return s  
            if s[-3:] == 'ing': return s + 'ly'  
            return s + 'ing'  
  
        print add_string('string')  
        print add_string('abc')
```

stringly  
abcing

Ex2. Write a Python program to remove the nth index character from a nonempty string.  
Expect result

```
>>> remove_char('Python', 0)  
    'ython'  
>>> remove_char('Python', 3)  
    'Pyton'  
>>> remove_char('Python', 5)  
    'Pytho'
```

```
In [3]: # Solution
def remove_char(s, index):
    return s[:index] + s[index+1:]

remove_char('Python', 0)
```

```
Out[3]: 'ython'
```

Ex3. Write a Python function to insert a string in the middle of a string. Using `remove_char()` function in Ex2 to remove the middle char.

```
>>> insert_middle('{{}}', 'JVN')
"{{JVN}}"
```

```
>>> insert_middle('[x]', 'JVN')
"[[JVN]]"
```

```
In [7]: # Solution
def insert_middle(s, sub_s):
    if len(s) % 2 == 1:
        s = remove_char(s, len(s)/2)

    middle_index = len(s) / 2
    return s[:middle_index] + sub_s + s[middle_index:]

insert_middle('[x]', 'JVN')
```

```
Out[7]: '[[JVN]]'
```

Ex4. Write a Python program to count the occurrences of each word in a given sentence

```
>>> word_count('the quick brown fox jumps over the lazy dog')
{'the': 2, 'jumps': 1, 'brown': 1, 'lazy': 1, 'fox': 1, 'over': 1, 'quick': 1, 'dog': 1}
```

```
In [8]: def word_count(s):
        s = s.split()
        wc = {}
        for word in s:
            wc[word] = wc.get(word, 0) + 1

        return wc

word_count('the quick brown fox jumps over the lazy dog')
```

```
Out[8]: {'brown': 1,
        'dog': 1,
        'fox': 1,
        'jumps': 1,
        'lazy': 1,
        'over': 1,
        'quick': 1,
        'the': 2}
```

Ex5. Write a Python program to get the domain of URL

```
>>> get_domain("https://www.kaggle.com/wiki/Tutorials")
"www.kaggle.com"
```

```
>>> get_domain("http://www.kaggle.com/wiki/Tutorials")
"www.kaggle.com"
```

```
>>> get_domain("jvn.edu.vn/xyz")
"jvn.edu.vn"
```

Extras: \* Using another library \* Using regex to detect domain. Library re

```
In [10]: # Solution 1
def get_domain(url = ''):
    url = url.replace('http://', '').replace('https://', '')
    return url.split('/')[0]

print get_domain("https://www.kaggle.com/wiki/Tutorials")

www.kaggle.com
```

```
In [12]: # Solution extra 1
from urlparse import urlparse

def get_domain(url = ''):
    o = urlparse(url)
    return o.netloc

print get_domain("https://www.kaggle.com/wiki/Tutorials")

www.kaggle.com
```

```
In [25]: # Solution extra 2
import re

def get_domain(url = ''):
    domain_patern = re.compile(r"^(https?:\\\/\\\/)?([a-z0-9]+\.[a-z]+\.[a-z]+?)\\\/")
    o = domain_patern.match(url)
    return o.group(2)

print get_domain("http://www.kaggle.com/wiki/Tutorials")
print get_domain("www.kaggle.com/wiki/Tutorials")

www.kaggle.com
www.kaggle.com
```

Read more: Regular expression (<https://docs.python.org/2/howto/regex.html>)

## 2 Import libraries

The outermost statements in a Python file, or "module", do its one-time setup — those statements run from top to bottom the first time the module is imported somewhere, setting up its variables and functions.

```
In [14]: import math

        print math.sqrt(4)
```

2.0

```
In [15]: from math import sqrt

        print sqrt(4)
```

2.0

```
In [16]: from math import *

        print sqrt(4)
```

2.0

```
In [17]: import math as m

        print m.sqrt(4)
```

2.0

## 3 Conditionals

Ex: Input a number then assign to variable **n**. If **n** larger than 10, write "larger than 10", else write "doesn't larger than 10"

### 3.1 Boolean expressions

A boolean expression is an expression that is either true or false. One way to write a boolean expression is to use the operator `==`, which compares two values and produces a boolean value:

```
>>> 5 == 5
True
>>> 5 == 6
False
```



In the first statement, the two operands are equal, so the value of the expression is True; in the second statement, 5 is not equal to 6, so we get False. True and False are special values that are built into Python.

The == operator is one of the comparison operators; the others are:

```
x != y      # x is not equal to y
x > y      # x is greater than y
x < y      # x is less than y
x >= y     # x is greater than or equal to y
x <= y     # x is less than or equal to y
```

### 3.2 Logical operators

There are three **logical operators**: and, or, and not. The semantics (meaning) of these operators is similar to their meaning in English. For example,  $x > 0$  and  $x < 10$  is true only if  $x$  is greater than 0 and less than 10.

$n\%2 == 0$  or  $n\%3 == 0$  is true if either of the conditions is true, that is, if the number is divisible by 2 or 3.

Finally, the not operator negates a boolean expression, so  $\text{not}(x > y)$  is true if  $(x > y)$  is false, that is, if  $x$  is less than or equal to  $y$ .

Strictly speaking, the operands of the logical operators should be boolean expressions, but Python is not very strict. Any nonzero number is interpreted as "true."

```
>>> x = 5
>>> x and 1
1
>>> y = 0
>>> y and 1
0
```

In general, this sort of thing is not considered good style. If you want to compare a value to zero, you should do it explicitly.

### 3.3 The "in" operator

The "in" operator could be used to check if a specified object exists within an iterable object container, such as a list:

```
In [18]: name = "John"
        name in ["John", "Rick"]

Out[18]: True

In [19]: "Duyet" in ["John", "Rick"]

Out[19]: False

In [30]: 'd' in 'dsdfdsfd'

Out[30]: True
```

## 3.4 IF statements

### 3.4.1 if-else

The basic if-else statement in Python is as follows:

```
if <expression>:
    #write statement here
else:
    #write statement here
```

```
In [31]: mark = int(input("What is your mark: "))
        if mark >= 50:
            print("Pass")
        else:
            print("Fail")
```

What is your mark: 5  
Fail

### 3.4.2 elif

```
In [21]: month = int(input("What is the current month ? "))
        if month <=0 or month >12:
            print "Invalid month number !!!"
        elif month < 4:
            print "Spring"
        elif month < 7:
            print "Summer"
        elif month < 10:
            print "Fall"
        elif month < 13:
            print "Winter"
```

What is the current month ? 3  
Spring

Bonus. Write a Python program to print alphabet pattern 'E' (using if, else)

```
*****
*
*
****
*
*
*****
```

## 4 Loops

### 4.1 For loops

For loops iterate over a given sequence. Here is an example:

```
In [22]: primes = [2, 3, 5, 7]
        for prime in primes:
            print(prime)
```

```
2
3
5
7
```

For loops can iterate over a sequence of numbers using the "range" and "xrange" functions.

```
In [23]: r = range(5)
        print r
```

```
[0, 1, 2, 3, 4]
```

```
In [24]: for i in r:
        print i
```

```
0
1
2
3
4
```

```
In [25]: for i in r:
        print "I love you"
```

```
I love you
I love you
I love you
I love you
I love you
```

The **range()** function is one of Python's built in functions. It is used to indicate how many times the loop will be repeated.

The structure of the range function is **range(start, upto, step)** in which the arguments of range are used as follows:

- **start** and **step** are both optional.

- **upto** must always be there, it means "up to but not including" the value.
- **start**, **upto**, and **step** must all be integers

```
In [26]: range(3)
```

```
Out[26]: [0, 1, 2]
```

```
In [27]: range(3,7)
```

```
Out[27]: [3, 4, 5, 6]
```

```
In [28]: range(1, 10, 2)
```

```
Out[28]: [1, 3, 5, 7, 9]
```

```
In [29]: range(10, -1, -1)
```

```
Out[29]: [10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0]
```

Ex1. Write a Python program to calculate the sum and average of n integer numbers (input from the user). Input 0 to finish

```
Input number: 6
```

```
Input number: 7
```

```
Input number: 10
```

```
Input number: 0
```

```
Sum = 23
```

```
AVG = 7.666666666666667
```

```
In [2]: # Solution 1
```

```
    number = 1
```

```
    Sum = 0
```

```
    n = 0
```

```
    while number != 0:
```

```
        number = int(raw_input("Input number: "))
```

```
        Sum += number # Sum = Sum + n
```

```
        n += 1
```

```
    print "Sum = ", Sum
```

```
    print "Avg = ", float(Sum) / (n-1)
```

```
Input number: 6
```

```
Input number: 7
```

```
Input number: 10
```

```
Input number: 0
```

```
Sum = 23
```

```
Avg = 7.666666666667
```

```

In [3]: # Solution 2
Sum = 0
n = 0

while True:
    number = int(raw_input("Input number: "))
    if number == 0:
        break

    Sum += number
    n += 1

print "Sum = ", Sum
print "Avg = ", float(Sum) / n

```

```

Input number: 6
Input number: 7
Input number: 10
Input number: 0
Sum = 23
Avg = 7.666666666667

```

Ex2. Write a Python program to create the multiplication table (from 1 to 10) of a number.

```

Input a number: 6
6 x 1 = 6
6 x 2 = 12
6 x 3 = 18
6 x 4 = 24
6 x 5 = 30
6 x 6 = 36
6 x 7 = 42
6 x 8 = 48
6 x 9 = 54
6 x 10 = 60

```

```

In [5]: # Solution
n = int(input("Input a number: "))
for i in range(1, 11):
    print n, "x", i, "=", n*i

```

```

Input a number: 6
6 x 1 = 6
6 x 2 = 12
6 x 3 = 18
6 x 4 = 24
6 x 5 = 30
6 x 6 = 36

```

```
6 x 7 = 42
6 x 8 = 48
6 x 9 = 54
6 x 10 = 60
```

Ex3. Write a Python program to construct the following pattern, using a nested loop number.

```
n = 9
```

```
1
22
333
4444
55555
666666
7777777
88888888
999999999
```

```
In [8]: # Solution
        n = int(input("n ="))

        print "\n"
        for i in range(1, n+1):
            for j in range(1, i+1):
                print i,
            print
```

```
n =9
```

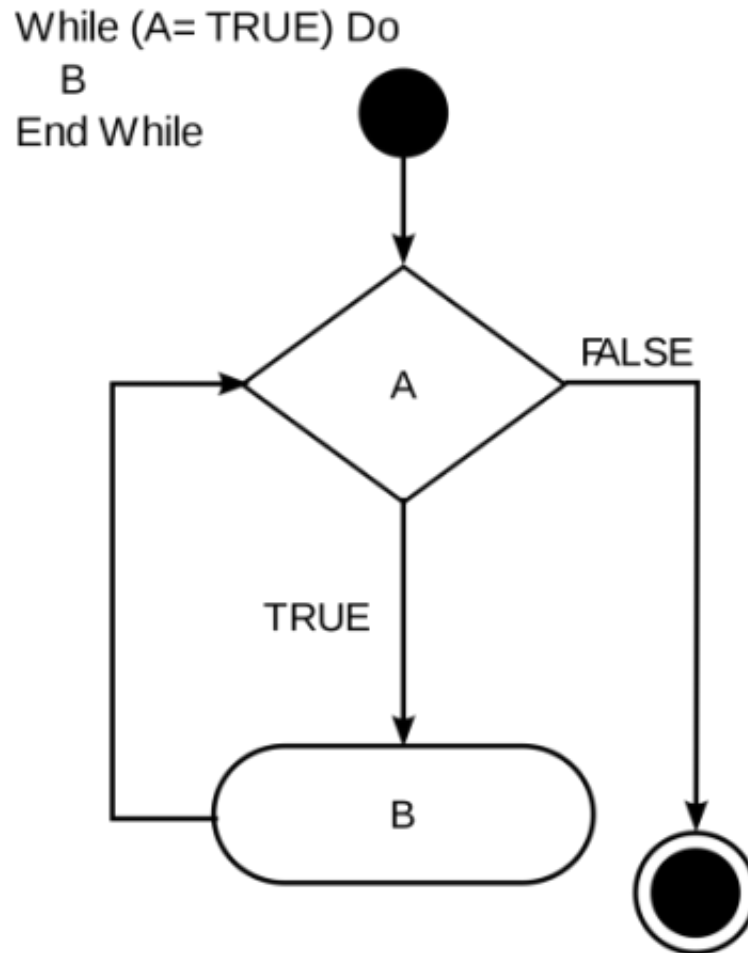
```
1
2 2
3 3 3
4 4 4 4
5 5 5 5 5
6 6 6 6 6 6
7 7 7 7 7 7 7
8 8 8 8 8 8 8 8
9 9 9 9 9 9 9 9 9
```

## 4.2 While loops

While loops are known as indefinite or conditional loops. They will keep iterating until certain conditions are met. There is no guarantee ahead of time regarding how many times the loop will iterate.

The while loop, like the if statement, includes a boolean expression that evaluates to true or false. The code inside the loop will be repeatedly executed until the boolean expression is no longer true.

This diagram shows the flow of control in a while loop:



```
In [30]: count = 0
        while count < 5:
            print(count)
            count += 1  # This is the same as count = count + 1
```

```
0
1
2
3
4
```

```
In [ ]: input_text = "run"
        while input_text != "exit":
```

```
input_text = raw_input("Please type 'exit' to quit this program: ")
print "Quit !!!"
```

### 4.3 "break" and "continue" statements

**break** is used to **exit** a for loop or a while loop, whereas **continue** is used to skip the current block, and return to the "for" or "while" statement. A few examples:

```
In [32]: count = 0
        while True:
            print(count)
            count += 1
            if count >= 5:
                break
```

0  
1  
2  
3  
4

```
In [33]: for x in range(10):
        # Check if x is even
        if x % 2 == 0:
            continue
        print(x)
```

1  
3  
5  
7  
9

### 4.4 "Infinite loops" and break

Sometimes you don't know it's time to end a loop until you get half way through the body. In that case you can write an infinite loop on purpose and then use the break statement to jump out of the loop.

### 4.5 Exercise

Ex1. Loop through and print out all even numbers from the numbers list in the same order they are received. Don't print any numbers that come after 237 in the sequence.

Hint: Using loop, continue and break

```
numbers = [
```



```

951, 402, 984, 651, 360, 69, 408, 319, 601, 485, 980, 507, 725, 547, 544,
615, 83, 165, 141, 501, 263, 617, 865, 575, 219, 390, 984, 592, 236, 105, 942, 941,
386, 462, 47, 418, 907, 344, 236, 375, 823, 566, 597, 978, 328, 615, 953, 345,
399, 162, 758, 219, 918, 237, 412, 566, 826, 248, 866, 950, 626, 949, 687, 217,
815, 67, 104, 58, 512, 24, 892, 894, 767, 553, 81, 379, 843, 831, 445, 742, 717,
958, 609, 842, 451, 688, 753, 854, 685, 93, 857, 440, 380, 126, 721, 328, 753, 470,
743, 527
]

# your code goes here

In [10]: # Solution
numbers = [
    951, 402, 984, 651, 360, 69, 408, 319, 601, 485, 980, 507, 725, 547, 544,
    615, 83, 165, 141, 501, 263, 617, 865, 575, 219, 390, 984, 592, 236, 105, 942, 941,
    386, 462, 47, 418, 907, 344, 236, 375, 823, 566, 597, 978, 328, 615, 953, 345,
    399, 162, 758, 219, 918, 237, 412, 566, 826, 248, 866, 950, 626, 949, 687, 217,
    815, 67, 104, 58, 512, 24, 892, 894, 767, 553, 81, 379, 843, 831, 445, 742, 717,
    958, 609, 842, 451, 688, 753, 854, 685, 93, 857, 440, 380, 126, 721, 328, 753, 470,
    743, 527
]

for n in numbers:
    if n == 237:
        break
    if n % 2 == 0:
        continue
    print n,

```

951 651 69 319 601 485 507 725 547 615 83 165 141 501 263 617 865 575 219 105 941 47 907 375 823

## Ex2. Faces on Money

It is common for images of a country's previous leaders, or other individuals of historical significance, to appear on its money. The individuals that appear on banknotes in the United States are listed in the table below.

Individual	Amount
George Washington	\$1
Thomas Jefferson	\$2
Abraham Lincoln	\$5
Alexander Hamilton	\$10
Andrew Jackson	\$20
Ulysses S. Grant	\$50
Benjamin Frankli	\$100

Write a program that begins by reading the denomination of a banknote from the user. Then

your program should display the name of the individual that appears on the banknote of the entered amount. An appropriate error message should be displayed if no such note exists.

In [1]: *# Solution*

```
banknotes = {
    1: 'George Washington',
    2: 'Thomas Jefferson',
    5: 'Abraham Lincoln',
    10: 'Alexander Hamilton',
    20: 'Andrew Jackson',
    50: 'Ulysses S. Grant',
    100: 'Benjamin Frankli'
}

# reading the denomination of a banknote
amount = int(input("Please enter amount: "))

while amount > 0:
    if amount / 100 > 0:
        print (amount / 100), ' x ', banknotes[100], '($100)'
        amount = amount % 100
    if amount / 50 > 0:
        print (amount / 50), ' x ', banknotes[50], '($50)'
        amount = amount % 50
    if amount / 20 > 0:
        print (amount / 20), ' x ', banknotes[20], '($20)'
        amount = amount % 20
    if amount / 10 > 0:
        print (amount / 10), ' x ', banknotes[10], '($10)'
        amount = amount % 10
    if amount / 5 > 0:
        print (amount / 5), ' x ', banknotes[5], '($5)'
        amount = amount % 5
    if amount / 2 > 0:
        print (amount / 2), ' x ', banknotes[2], '($2)'
        amount = amount % 2
    if amount / 1 > 0:
        print (amount / 1), ' x ', banknotes[1], '($1)'
        amount = amount % 1
```

```
Please enter amount: 345
3 x Benjamin Frankli ($100)
2 x Andrew Jackson ($20)
1 x Abraham Lincoln ($5)
```

In [16]: *# Solution for shorter*

```

banknotes = {
    1: 'George Washington',
    2: 'Thomas Jefferson',
    5: 'Abraham Lincoln',
    10: 'Alexander Hamilton',
    20: 'Andrew Jackson',
    50: 'Ulysses S. Grant',
    100: 'Benjamin Frankli'
}

# reading the denomination of a banknote
amount = int(input("Please enter amount: "))

while amount > 0:
    for i in sorted(banknotes.keys(), reverse=True):
        if amount / i > 0:
            print (amount / i), ' x ', banknotes[i], '($'+str(i)+')'
            amount = amount % i
            break

```

```

Please enter amount: 456
4 x Benjamin Frankli ($100)
1 x Ulysses S. Grant ($50)
1 x Abraham Lincoln ($5)
1 x George Washington ($1)

```

Ex3. (Bonus) Implement insertion sort in Python.

```
insertion_sort([3,4,6,7,1])
```

Ex4. (Bonus) Implement selection sort in Python.

```
selection_sort([3,4,6,7,1])
```

## 5 References

- Class and Object: [https://www.learnpython.org/en/Classes\\_and\\_Objects](https://www.learnpython.org/en/Classes_and_Objects)
- Learn Python - Loop: <https://www.learnpython.org/en/Loops>
- <http://www.afterhoursprogramming.com/tutorial/Python/If-Statement/>
- <http://www.afterhoursprogramming.com/tutorial/Python/Classes/>