

# Python数据科学速查表

## Python基础

学习更多Python数据科学[www.datacamp.com](http://www.datacamp.com)



### 变量和数据类型

#### 变量赋值

```
>>> x=5
>>> x
5
```

#### 变量计算

>>> x+2	和
>>> x	
5	
>>> x-2	差
>>> x*2	积
>>> x**2	乘方
>>> x%2	余数
>>> x/float(2)	除
2.5	

#### 类型和类型转换

str()	'5', '3.45', 'True'	转字符串
int()	5, 3, 1	转整型
float()	5.0, 1.0	转浮点型
bool()	True, True, True	转布尔型

#### 帮助

```
>>> help(str)
```

### 字符串

```
>>> my_string = 'thisStringIsAwesome'
>>> my_string
'thisStringIsAwesome'
```

#### 字符串操作

```
>>> my_string * 2
'thisStringIsAwesomethisStringIsAwesome'
>>> my_string + 'Innit'
'thisStringIsAwesomeInnit'
>>> 'm' in my_string
True
```

### 列表

Also see NumPy Arrays

```
>>> a = 'is'
>>> b = 'nice'
>>> my_list = ['my', 'list', a, b]
>>> my_list2 = [[4,5,6,7], [3,4,5,6]]
```

#### 选择列表元素

Index starts at 0

##### Subset

```
>>> my_list[1]
>>> my_list[-3]
```

选择索引为1的元素  
倒数第3个元素

##### Slice

```
>>> my_list[1:3]
>>> my_list[1:]
>>> my_list[:3]
>>> my_list[:]
```

选择索引1和2的元素  
选择从索引1开始的元素  
选择索引3以前元素，不含3  
复制列表

##### Subset Lists of Lists

```
>>> my_list2[1][0]
>>> my_list2[1][:2]
```

my\_list[list][itemOfList]

#### 列表操作

```
>>> my_list + my_list
['my', 'list', 'is', 'nice', 'my', 'list', 'is', 'nice']
>>> my_list * 2
['my', 'list', 'is', 'nice', 'my', 'list', 'is', 'nice']
>>> my_list2 > 4
True
```

#### 列表方法

```
>>> my_list.index(a)
>>> my_list.count(a)
>>> my_list.append('!')
>>> my_list.remove('!')
>>> del(my_list[0:1])
>>> my_list.reverse()
>>> my_list.extend('!')
>>> my_list.pop(-1)
>>> my_list.insert(0, '!')
>>> my_list.sort()
```

获得元素索引  
计算元素个数  
追加项  
移除项  
删除  
翻转列表  
扩展项  
取列表最后一项  
索引0处插入  
列表排序

#### 字符串操作

Index starts at 0

```
>>> my_string[3]
>>> my_string[4:9]
```

#### 字符串方法

```
>>> my_string.upper()
>>> my_string.lower()
>>> my_string.count('w')
>>> my_string.replace('e', 'i')
>>> my_string.strip()
```

字符串转大写  
字符串转小写  
字符串中字符计数  
替换字符，'e' -> 'i'  
去首尾空格

### 库

#### Import libraries

```
>>> import numpy
>>> import numpy as np
Selective import
>>> from math import pi
```

pandas  
Data analysis

Machine learning

NumPy  
Scientific computing

matplotlib  
2Dplotting

### 安装Python

ANACONDA  
Leading open data science platform  
powered by Python

spyder  
Free IDE that is included  
with Anaconda

jupyter  
Create and share  
documents with live code,  
visualizations, text, ...

### Numpy 数组

Also see Lists

```
>>> my_list = [1, 2, 3, 4]
>>> my_array = np.array(my_list)
>>> my_2darray = np.array([[1,2,3], [4,5,6]])
```

#### 选择Numpy数组元素

Index starts at 0

##### Subset

```
>>> my_array[1]
2
```

选择索引为1的元素

##### Slice

```
>>> my_array[0:2]
array([1, 2])
```

选择索引为[0,2)的元素

##### Subset 2D Numpy arrays

```
>>> my_2darray[:,0]
array([1, 4])
```

my\_2darray[rows, columns]

#### Numpy数组操作

```
>>> my_array > 3
array([False, False, False,  True], dtype=bool)
>>> my_array * 2
array([2, 4, 6, 8])
>>> my_array + np.array([5, 6, 7, 8])
array([6, 8, 10, 12])
```

#### Numpy数组功能

```
>>> my_array.shape
>>> np.append(other_array)
>>> np.insert(my_array, 1, 5)
>>> np.delete(my_array, [1])
>>> np.mean(my_array)
>>> np.median(my_array)
>>> my_array.corrcoef()
>>> np.std(my_array)
```

数组维数  
追加数组项  
插入数组项  
删除数组项  
数组均值  
数组中位数  
相关系数  
标准差

DataCamp

Interactively