



# Python 数据科学 速查表

## Python 基础

天善智能 商业智能与大数据社区 [www.hellobi.com](http://www.hellobi.com)



### 变量与数据类型

#### 变量赋值

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

#### 变量计算

<code>&gt;&gt;&gt; x+2</code>	加
7	
<code>&gt;&gt;&gt; x-2</code>	减
3	
<code>&gt;&gt;&gt; x*2</code>	乘
10	
<code>&gt;&gt;&gt; x**2</code>	幂
25	
<code>&gt;&gt;&gt; x%2</code>	取余
1	
<code>&gt;&gt;&gt; x/float(2)</code>	除
2.5	

#### 类型与类型转换

<code>str()</code>	'5', '3.45', 'True'	转为字符串
<code>int()</code>	5, 3, 1	转为整数
<code>float()</code>	5.0, 1.0	转为浮点数
<code>bool()</code>	True, True, True	转为布尔值

#### 调用帮助

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

### 字符串

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

#### 字符串运算

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

### 列表

```
>>> 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

索引始于0

##### 子集

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

##### 切片

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

##### 子集列表的列表

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

#### 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()
```

获取某值的索引  
统计某值出现的次数  
追加某值  
移除某值  
反转列表  
添加某值  
移除某值  
插入某值  
列表排序

### 参阅 Numpy 数组

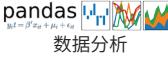
### Python库

#### 导入库

```
>>> import numpy
>>> import numpy as np
```

#### 导入指定功能

```
>>> from math import pi
```

pandas   
数据分析

learn   
机器学习

NumPy   
科学计算

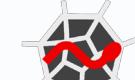
matplotlib   
二维视图

### 安装 Python



**ANACONDA**

Python 首选开源数据科学平台



Anaconda  
内置的免费IDE



创建包含代码、可视图  
与文本的文档

### Numpy 数组

参阅 列表

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

#### 选取 Numpy 数组的值

索引始于0

##### 子集

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

##### 切片

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

##### 二维 Numpy 数组的子集

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

选择索引1对应的值

选择索引0和1对应的值

my\_2darray[rows, columns]

#### Numpy 数组运算

```
>>> 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 数组函数

<code>&gt;&gt;&gt; my_array.shape</code>	获取数组形状
<code>&gt;&gt;&gt; np.append(other_array)</code>	追加数据
<code>&gt;&gt;&gt; np.insert(my_array, 1, 5)</code>	插入数据
<code>&gt;&gt;&gt; np.delete(my_array, [1])</code>	删除数据
<code>&gt;&gt;&gt; np.mean(my_array)</code>	平均值
<code>&gt;&gt;&gt; np.median(my_array)</code>	中位数
<code>&gt;&gt;&gt; my_array.corrcoef()</code>	相关系数
<code>&gt;&gt;&gt; np.std(my_array)</code>	标准差

获取数组形状  
追加数据  
插入数据  
删除数据  
平均值  
中位数  
相关系数  
标准差

### 原文作者

DataCamp

Learn Python for Data Science Interactively!

