



AniPython

pandas简介

pandas是什么？

pandas是一个开源的、BSD许可的库
为Python编程语言提供高性能、易于使用的数据结构和数据分析工具
· · · · ·

列表 / numpy / pandas

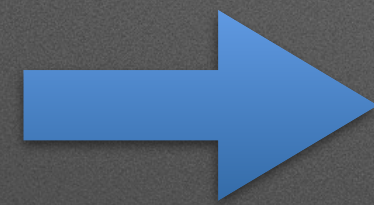
~~list~~

列表 / numpy / pandas

列表 / numpy / pandas

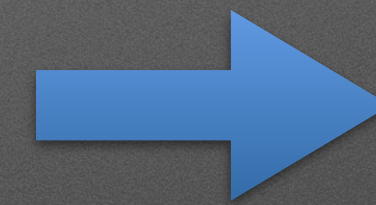
[[1, 2],
[3, 4]]

<list>



[[1 2]
[3 4]]

<numpy.array>



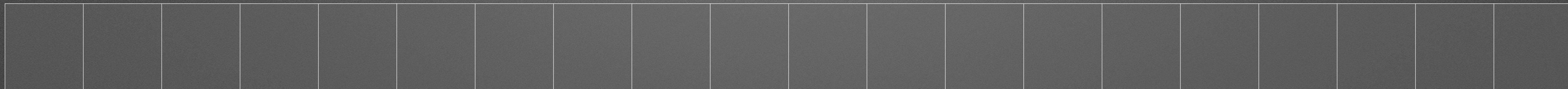
x
y

<pandas.DataFrame>

A	B
1	2
3	4

高性能

内存



list: 1 2 3

numpy/pandas: 1 2 3

高性能

内存

	2	1		3															
--	---	---	--	---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

list:

numpy/pandas: 1 2 3

高性能

内存

	2	1		3								1	2	3					
--	---	---	--	---	--	--	--	--	--	--	--	---	---	---	--	--	--	--	--

list:

numpy/pandas:

高性能

内存

	2	1		3		10	30		20			1	2	3		10	20	30	
--	---	---	--	---	--	----	----	--	----	--	--	---	---	---	--	----	----	----	--

list: *10

numpy/pandas: *10

易用

	0	1	2
0	180	300	100
1	120	250	200
2	80	199	300

numpy: <ndarray>

易用

		0	1	2
		成本	售价	销量
0	羽绒服	180	300	100
1	家居服	120	250	200
2	牛仔裤	80	199	300

columns
列标签
列索引
列名
特征

index
行标签
行索引
行名
样本ID

pandas: <DataFrame>

易用

	成本	售价	销量
羽绒服	180	300	100
家居服	120	250	200
牛仔裤	80	199	300

获取"家居服"的"销量"?

```
sales = df.loc["家居服", "销量"] # 200
```


易用

	成本	售价	销量	总销售额
羽绒服	180	300	100	30000
家居服	120	250	200	50000
牛仔裤	80	199	300	59700

获取"家居服"的"销量"?

```
sales = df.loc["家居服", "销量"] # 200
```

增加一列"总销售额"

```
df["总销售额"] = df["售价"] * df["销量"]
```