

pandas中的数据选取操作

```
In [1]: import pandas as pd
```

```
In [3]: data = {  
        'name': ['张三', '李四', '王五', '赵六'],  
        'age': [19, 20, 18, 21],  
        'height': [1.68, 1.72, 1.64, 1.55]  
    }  
df = pd.DataFrame(data, columns=['name', 'age', 'height'])  
df
```

Out[3]:

| | name | age | height |
|---|------|-----|--------|
| 0 | 张三 | 19 | 1.68 |
| 1 | 李四 | 20 | 1.72 |
| 2 | 王五 | 18 | 1.64 |
| 3 | 赵六 | 21 | 1.55 |

列操作

获取一列或多列数据

```
In [4]: df['name']
```

```
Out[4]: 0    张三  
        1    李四  
        2    王五  
        3    赵六  
        Name: name, dtype: object
```

```
In [5]: df.name
```

```
Out[5]: 0    张三  
        1    李四  
        2    王五  
        3    赵六  
        Name: name, dtype: object
```

```
In [7]: df[['name']]
```

Out[7]:

| | name |
|---|------|
| 0 | 张三 |
| 1 | 李四 |
| 2 | 王五 |
| 3 | 赵六 |

```
In [8]: df[['name', 'age']]
```

Out[8]:

| | name | age |
|---|------|-----|
| 0 | 张三 | 19 |
| 1 | 李四 | 20 |
| 2 | 王五 | 18 |
| 3 | 赵六 | 21 |

修改一列数据 **copy** & **not copy**

```
In [9]: names = df.name  
names
```

Out[9]:

| | |
|---|----|
| 0 | 张三 |
| 1 | 李四 |
| 2 | 王五 |
| 3 | 赵六 |

Name: name, dtype: object

不使用copy的数据,如果对其进行修改, 原数据也被改动

```
In [10]: names[0] = '田七'
```

```
/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
```

```
"""Entry point for launching an IPython kernel.
```

```
In [11]: names
```

```
Out[11]: 0    田七  
1    李四  
2    王五  
3    赵六  
Name: name, dtype: object
```

```
In [12]: df
```

```
Out[12]:
```

| | name | age | height |
|---|------|-----|--------|
| 0 | 田七 | 19 | 1.68 |
| 1 | 李四 | 20 | 1.72 |
| 2 | 王五 | 18 | 1.64 |
| 3 | 赵六 | 21 | 1.55 |

使用了copy方法，改动数据就不会对原数据造成影响

```
In [13]: names = df.name.copy()  
names[0] = '周八'  
names
```

```
Out[13]: 0    周八  
1    李四  
2    王五  
3    赵六  
Name: name, dtype: object
```

```
In [14]: df
```

```
Out[14]:
```

| | name | age | height |
|---|------|-----|--------|
| 0 | 田七 | 19 | 1.68 |
| 1 | 李四 | 20 | 1.72 |
| 2 | 王五 | 18 | 1.64 |
| 3 | 赵六 | 21 | 1.55 |

```
In [15]: df.columns
```

```
Out[15]: Index(['name', 'age', 'height'], dtype='object')
```

```
In [16]: df[df.columns[1:3]]
```

Out[16]:

| | age | height |
|---|-----|--------|
| 0 | 19 | 1.68 |
| 1 | 20 | 1.72 |
| 2 | 18 | 1.64 |
| 3 | 21 | 1.55 |

添加一列数据

```
In [17]: import datetime
df['year'] = datetime.datetime.now().year - df.age
df
```

Out[17]:

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 2 | 王五 | 18 | 1.64 | 2000 |
| 3 | 赵六 | 21 | 1.55 | 1997 |

删除一列数据

删除一列数据时，不会对原数据造成影响，如果要使用删除后的数据，需要单独接收

```
In [18]: df.drop('year', axis=1)
```

Out[18]:

| | name | age | height |
|---|------|-----|--------|
| 0 | 田七 | 19 | 1.68 |
| 1 | 李四 | 20 | 1.72 |
| 2 | 王五 | 18 | 1.64 |
| 3 | 赵六 | 21 | 1.55 |

In [19]: df

Out[19]:

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 2 | 王五 | 18 | 1.64 | 2000 |
| 3 | 赵六 | 21 | 1.55 | 1997 |

In [20]: df.drop(['height', 'year'], axis=1)

Out[20]:

| | name | age |
|---|------|-----|
| 0 | 田七 | 19 |
| 1 | 李四 | 20 |
| 2 | 王五 | 18 |
| 3 | 赵六 | 21 |

In [21]: df

Out[21]:

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 2 | 王五 | 18 | 1.64 | 2000 |
| 3 | 赵六 | 21 | 1.55 | 1997 |

In [22]: df.drop(df.columns[1:2], axis=1)

Out[22]:

| | name | height |
|---|------|--------|
| 0 | 田七 | 1.68 |
| 1 | 李四 | 1.72 |
| 2 | 王五 | 1.64 |
| 3 | 赵六 | 1.55 |

行操作

获取一行或多行数据

```
In [24]: df
```

```
Out[24]:
```

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 2 | 王五 | 18 | 1.64 | 2000 |
| 3 | 赵六 | 21 | 1.55 | 1997 |

使用 loc[] 其中填的是index的名字

```
In [26]: df.loc[1]
```

```
Out[26]: name      李四  
age         20  
height     1.72  
year       1998  
Name: 1, dtype: object
```

```
In [27]: type(df.loc[1])
```

```
Out[27]: pandas.core.series.Series
```

```
In [30]: df.loc[[1]]
```

```
Out[30]:
```

| | name | age | height | year |
|---|------|-----|--------|------|
| 1 | 李四 | 20 | 1.72 | 1998 |

```
In [31]: df.loc[[1, 3]]
```

```
Out[31]:
```

| | name | age | height | year |
|---|------|-----|--------|------|
| 1 | 李四 | 20 | 1.72 | 1998 |
| 3 | 赵六 | 21 | 1.55 | 1997 |

```
In [32]: df.index
```

```
Out[32]: RangeIndex(start=0, stop=4, step=1)
```

```
In [33]: df.index[-2:]
```

```
Out[33]: RangeIndex(start=2, stop=4, step=1)
```

```
In [34]: df.loc[df.index[-2:]]
```

```
Out[34]:
```

| | name | age | height | year |
|---|------|-----|--------|------|
| 2 | 王五 | 18 | 1.64 | 2000 |
| 3 | 赵六 | 21 | 1.55 | 1997 |

```
In [37]: df[1:3]
```

```
Out[37]:
```

| | name | age | height | year |
|---|------|-----|--------|------|
| 1 | 李四 | 20 | 1.72 | 1998 |
| 2 | 王五 | 18 | 1.64 | 2000 |

```
In [38]: df[-2:]
```

```
Out[38]:
```

| | name | age | height | year |
|---|------|-----|--------|------|
| 2 | 王五 | 18 | 1.64 | 2000 |
| 3 | 赵六 | 21 | 1.55 | 1997 |

获取部分行部分列

```
In [35]: df.loc[df.index[-2:], ['name', 'year']]
```

```
Out[35]:
```

| | name | year |
|---|------|------|
| 2 | 王五 | 2000 |
| 3 | 赵六 | 1997 |

```
In [36]: df
```

```
Out[36]:
```

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 2 | 王五 | 18 | 1.64 | 2000 |
| 3 | 赵六 | 21 | 1.55 | 1997 |

```
In [39]: df.shape
```

```
Out[39]: (4, 4)
```

添加一行数据

```
In [41]: df.loc[df.shape[0]] = {'name': 'jack', 'age': 23, 'height': 1.86, 'year': 1995}
df
```

Out[41]:

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 2 | 王五 | 18 | 1.64 | 2000 |
| 3 | 赵六 | 21 | 1.55 | 1997 |
| 4 | jack | 23 | 1.86 | 1995 |

```
In [43]: df2 = df.drop(2)
df2
```

Out[43]:

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 3 | 赵六 | 21 | 1.55 | 1997 |
| 4 | jack | 23 | 1.86 | 1995 |

重置index

```
In [44]: df2.index = range(df2.shape[0])
df2
```

Out[44]:

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 2 | 赵六 | 21 | 1.55 | 1997 |
| 3 | jack | 23 | 1.86 | 1995 |


```
In [45]: df2 = df.drop(2)
df2
```

Out[45]:

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 3 | 赵六 | 21 | 1.55 | 1997 |
| 4 | jack | 23 | 1.86 | 1995 |

使用iloc(), 其中的值为index的从0开始的序号, 比如下面案例中的2, 是指index序号为2, 并非index为2

```
In [46]: df2.iloc[2]          # integer location
```

Out[46]:

| | |
|--------|------|
| name | 赵六 |
| age | 21 |
| height | 1.55 |
| year | 1997 |

Name: 3, dtype: object

```
In [47]: df2.index = list('ABCD')
df2
```

Out[47]:

| | name | age | height | year |
|---|------|-----|--------|------|
| A | 田七 | 19 | 1.68 | 1999 |
| B | 李四 | 20 | 1.72 | 1998 |
| C | 赵六 | 21 | 1.55 | 1997 |
| D | jack | 23 | 1.86 | 1995 |

```
In [49]: df2.iloc[1:3]
```

Out[49]:

| | name | age | height | year |
|---|------|-----|--------|------|
| B | 李四 | 20 | 1.72 | 1998 |
| C | 赵六 | 21 | 1.55 | 1997 |

注意: 区别 iloc[1:3] 与 iloc[1,3] 的区别

```
In [51]: df2.iloc[1,3]      # 获取第1行第3列的数据    行和列从0开始计数
```

Out[51]: 1998

获取某一行某一列的某个数据： `iloc[a, b]` 与 `iat[a, b]`

```
In [52]: df2
```

```
Out[52]:
```

| | name | age | height | year |
|---|------|-----|--------|------|
| A | 田七 | 19 | 1.68 | 1999 |
| B | 李四 | 20 | 1.72 | 1998 |
| C | 赵六 | 21 | 1.55 | 1997 |
| D | jack | 23 | 1.86 | 1995 |

```
In [54]: df2.iat[1, 1]      # 同 df2.iloc[1, 1]
```

```
Out[54]: 20
```

数据筛选

```
In [55]: df
```

```
Out[55]:
```

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 2 | 王五 | 18 | 1.64 | 2000 |
| 3 | 赵六 | 21 | 1.55 | 1997 |
| 4 | jack | 23 | 1.86 | 1995 |

```
In [56]: df['height'] >= 1.65
```

```
Out[56]: 0      True
         1      True
         2     False
         3     False
         4      True
         Name: height, dtype: bool
```

```
In [57]: df[df['height'] >= 1.65]
```

Out[57]:

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 4 | jack | 23 | 1.86 | 1995 |

```
In [58]: df[(df['height'] >= 1.65) & (df['age']<=20)]
```

Out[58]:

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |

```
In [59]: df.query('height>=1.65 and age<=20')      # 与SQL语句类似
```

Out[59]:

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |

```
In [61]: df.query('height>=1.65 and age<=20 or name=="jack"')
```

Out[61]:

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 4 | jack | 23 | 1.86 | 1995 |

```
In [62]: age = 20
df.query('age<=@age')
```

Out[62]:

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 2 | 王五 | 18 | 1.64 | 2000 |

```
In [63]: df['age'].isin([18, 19])
```

```
Out[63]: 0      True
          1     False
          2      True
          3     False
          4     False
          Name: age, dtype: bool
```

```
In [64]: df[df['age'].isin([18, 19])]
```

```
Out[64]:
```

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 2 | 王五 | 18 | 1.64 | 2000 |

```
In [65]: df
```

```
Out[65]:
```

| | name | age | height | year |
|---|------|-----|--------|------|
| 0 | 田七 | 19 | 1.68 | 1999 |
| 1 | 李四 | 20 | 1.72 | 1998 |
| 2 | 王五 | 18 | 1.64 | 2000 |
| 3 | 赵六 | 21 | 1.55 | 1997 |
| 4 | jack | 23 | 1.86 | 1995 |

数据转置

```
In [66]: df.T
```

```
Out[66]:
```

| | 0 | 1 | 2 | 3 | 4 |
|--------|------|------|------|------|------|
| name | 田七 | 李四 | 王五 | 赵六 | jack |
| age | 19 | 20 | 18 | 21 | 23 |
| height | 1.68 | 1.72 | 1.64 | 1.55 | 1.86 |
| year | 1999 | 1998 | 2000 | 1997 | 1995 |