blog.csdn.net

(28条消息)【python】详解pandas.DataFrame.loc函数_brucewong0516的博客-CSDN博客_dataframe loc

7-8 minutes

官方函数

```
DataFrame.loc

Access a group of rows and columns by label(s) or a boolean array.

.loc[] is primarily label based, but may also be used with a boolean array.

# 可以使用label值，但是也可以使用布尔值
```

- 1
- 2
- 3
- 4

Allowed inputs are: # 可以接受单个的label，多个label的列表，多个label的切片

•A single label, e.g. 5 or 'a', (note that 5 is interpreted as a label of the index, and never as an integer position along the index). #这里的5不是数值指定的位置，而是label值

•A list or array of labels, e.g. ['a', 'b', 'c'].

•A slice object with labels, e.g. 'a':'f'.

Warning: #如果使用多个label的切片，那么切片的起始位置都是包含的

Note that contrary to usual python slices, both the start and the stop are included

•A boolean array of the same length as the axis being sliced, e.g. [True, False, True].

实例详解

一、选择数值

1、生成df

```
df = pd.DataFrame([[1, 2], [4, 5], [7, 8]],
...      index=['cobra', 'viper', 'sidewinder'],
```

```
...          columns=['max_speed', 'shield'])
```

```
df
Out[15]:
```

	max_speed	shield
cobra	1	2
viper	4	5
sidewinder	7	8

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

2、Single label. 单个 row_label 返回的Series

```
 df.loc['viper']
Out[17]:
max_speed    4
shield       5
Name: viper, dtype: int64
```

- 1
- 2
- 3
- 4
- 5

2、List of labels. 列表 row_label 返回的DataFrame

```
df.loc[['cobra','viper']]
Out[20]:
```

```
        max_speed   shield

cobra          1        2

viper          4        5
```

- 1
- 2
- 3
- 4
- 5

3、Single label for row and column 同时选定行和列

```
df.loc['cobra', 'shield']

Out[24]: 2
```

- 1
- 2

4、Slice with labels for row and single label for column. As mentioned above, note that both the start and stop of the slice are included. 同时选定多个行和单个列，注意的是通过列表选定多个row label 时，首位均是选定的。

```
df.loc['cobra':'viper', 'max_speed']

Out[25]:

cobra     1

viper     4

Name: max_speed, dtype: int64
```

- 1
- 2
- 3
- 4
- 5

5、Boolean list with the same length as the row axis 布尔列表选择row label

布尔值列表是根据某个位置的True or False 来选定，如果某个位置的布尔值是True，则选定该row

```
df

Out[30]:
```

```
              max_speed   shield
cobra                 1        2
viper                 4        5
sidewinder            7        8
```

df.loc[[True]]

Out[31]:

```
         max_speed   shield
cobra            1        2
```

df.loc[[True,False]]

Out[32]:

```
         max_speed   shield
cobra            1        2
```

df.loc[[True,False,True]]

Out[33]:

```
              max_speed   shield
cobra                 1        2
sidewinder            7        8
```

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15

- 16
- 17
- 18
- 19
- 20
- 21
- 22

6、Conditional that returns a boolean Series 条件布尔值

```
df.loc[df['shield'] > 6]
Out[34]:
          max_speed   shield
sidewinder         7        8
```

- 1
- 2
- 3
- 4

7、Conditional that returns a boolean Series with column labels specified 条件布尔值和具体某列的数据

```
df.loc[df['shield'] > 6, ['max_speed']]
Out[35]:
          max_speed
sidewinder         7
```

- 1
- 2
- 3
- 4

8、Callable that returns a boolean Series 通过函数得到布尔结果选定数据

```
df
Out[37]:
          max_speed   shield
cobra              1        2
```

```
viper               4        5

sidewinder          7        8
```

```
df.loc[lambda df: df['shield'] == 8]
```

Out[38]:

```
            max_speed   shield

sidewinder          7        8
```

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11

二、赋值

1、Set value for all items matching the list of labels 根据某列表选定的row 及某列 column 赋值

```
df.loc[['viper', 'sidewinder'], ['shield']] = 50
```

```
df
```

Out[43]:

```
            max_speed   shield

cobra               1        2

viper               4       50

sidewinder          7       50
```

- 1
- 2
- 3
- 4
- 5

- 6
- 7
- 8

2、Set value for an entire row 将某行row的数据全部赋值

```
df.loc['cobra'] =10
```

```
df
Out[48]:
```

	max_speed	shield
cobra	10	10
viper	4	50
sidewinder	7	50

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

3、Set value for an entire column 将某列的数据完全赋值

```
df.loc[:, 'max_speed'] = 30
```

```
df
Out[50]:
```

	max_speed	shield
cobra	30	10
viper	30	50
sidewinder	30	50

- 1
- 2
- 3

- 4
- 5
- 6
- 7
- 8

4、Set value for rows **matching callable condition 条件选定rows赋值**

```
df.loc[df['shield'] > 35] = 0
```

```
df
```

Out[52]:

	max_speed	shield
cobra	30	10
viper	0	0
sidewinder	0	0

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

三、行索引是数值

```
df = pd.DataFrame([[1, 2], [4, 5], [7, 8]],
...        index=[7, 8, 9], columns=['max_speed', 'shield'])
```

```
df
```

Out[54]:

	max_speed	shield
7	1	2
8	4	5
9	7	8

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

通过 行 rows的切片的方式取多个：

```
df.loc[7:9]
Out[55]:
   max_speed   shield
7           1        2
8           4        5
9           7        8
```

- 1
- 2
- 3
- 4
- 5
- 6

四、多维索引

1、生成多维索引

```
tuples = [
...     ('cobra', 'mark i'), ('cobra', 'mark ii'),
...     ('sidewinder', 'mark i'), ('sidewinder', 'mark ii'),
...     ('viper', 'mark ii'), ('viper', 'mark iii')
... ]
index = pd.MultiIndex.from_tuples(tuples)
values = [[12, 2], [0, 4], [10, 20],
...          [1, 4], [7, 1], [16, 36]]
```

```
df = pd.DataFrame(values, columns=['max_speed', 'shield'], index=index)
```

df

Out[57]:

		max_speed	shield
cobra	mark i	12	2
	mark ii	0	4
sidewinder	mark i	10	20
	mark ii	1	4
viper	mark ii	7	1
	mark iii	16	36

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20

2、Single label. 传入的就是最外层的row label，返回DataFrame

```
df.loc['cobra']
Out[58]:
        max_speed   shield
mark i          12       2
mark ii          0       4
```

- 1
- 2
- 3
- 4
- 5

3、Single index tuple.传入的是索引元组，返回Series

```
df.loc[('cobra', 'mark ii')]
Out[59]:
max_speed    0
shield       4
Name: (cobra, mark ii), dtype: int64
```

- 1
- 2
- 3
- 4
- 5

4、Single label for row and column.如果传入的是row和column，和传入tuple是类似的，返回Series

```
df.loc['cobra', 'mark i']
Out[60]:
max_speed    12
shield        2
Name: (cobra, mark i), dtype: int64
```

- 1
- 2
- 3
- 4

- 5

5、Single tuple. Note using [[ ]] returns a DataFrame.传入一个数组，返回一个DataFrame

```
df.loc[[('cobra', 'mark ii')]]
Out[61]:
                 max_speed   shield
cobra mark ii           0        4
```

- 1
- 2
- 3
- 4

6、Single tuple for the index with a single label for the column 获取某个colum的某row的数据，需要左边传入多维索引的

tuple，然后再传入column

```
df.loc[('cobra', 'mark i'), 'shield']
Out[62]: 2
```

- 1
- 2

7、传入多维索引和单个索引的切片：

```
df.loc[('cobra', 'mark i'):'viper']
Out[63]:

                     max_speed   shield
cobra       mark i          12        2

            mark ii          0        4

sidewinder  mark i          10       20

            mark ii          1        4

viper       mark ii          7        1

            mark iii        16       36


df.loc[('cobra', 'mark i'):'sidewinder']
Out[64]:

                    max_speed   shield
```

```
cobra        mark i           12        2

             mark ii           0        4

sidewinder mark i           10       20

             mark ii           1        4
```

```
df.loc[('cobra', 'mark i'):('sidewinder','mark i')]
Out[65]:
                     max_speed   shield
cobra        mark i           12        2

             mark ii           0        4

sidewinder mark i           10       20
```

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24