pandas数据汇总

describe
count
mean
sum

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

In [2]:
```
data = [[1,  None],
            [4, 5],
            [None, None],
            [8, 9],
            [3, 4],
            [7, 7]]
df = pd.DataFrame(data, columns=['a', 'b'])
df
```

Out[2]:

	a	b
0	1.0	NaN
1	4.0	5.0
2	NaN	NaN
3	8.0	9.0
4	3.0	4.0
5	7.0	7.0

In [3]:
```
df.head()          # head()只显示前5条数据
```

Out[3]:

	a	b
0	1.0	NaN
1	4.0	5.0
2	NaN	NaN
3	8.0	9.0
4	3.0	4.0

In [4]:
```
df.head(2)
```

Out[4]:

	a	b
0	1.0	NaN
1	4.0	5.0

In [5]:
```
df.head(6)
```

Out[5]:

	a	b
0	1.0	NaN
1	4.0	5.0
2	NaN	NaN
3	8.0	9.0
4	3.0	4.0
5	7.0	7.0

In [6]:
```
df.tail()                    # tail()只显示最后5条数据
```

Out[6]:

	a	b
1	4.0	5.0
2	NaN	NaN
3	8.0	9.0
4	3.0	4.0
5	7.0	7.0

In [7]:
```
df.tail(3)
```

Out[7]:

	a	b
3	8.0	9.0
4	3.0	4.0
5	7.0	7.0

In [8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6 entries, 0 to 5
Data columns (total 2 columns):
a    5 non-null float64
b    4 non-null float64
dtypes: float64(2)
memory usage: 176.0 bytes
```

In [9]:
```
df.describe()        # 统计和计算时，都将None的数据排除在外的
```

Out[9]:

	a	b
count	5.000000	4.000000
mean	4.600000	6.250000
std	2.880972	2.217356
min	1.000000	4.000000
25%	3.000000	4.750000
50%	4.000000	6.000000
75%	7.000000	7.500000
max	8.000000	9.000000

In [10]:
```
df.count()
```

Out[10]:
```
a    5
b    4
dtype: int64
```

In [11]:
```
df.mean()        # 平均值
```

Out[11]:
```
a    4.60
b    6.25
dtype: float64
```

In [12]:
```
df.sum()
```

Out[12]:
```
a    23.0
b    25.0
dtype: float64
```

In [13]:
```
df.sum(axis=1)
```

Out[13]:
```
0     1.0
1     9.0
2     0.0
3    17.0
4     7.0
5    14.0
dtype: float64
```

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

Out[14]:

	a	b
0	1.0	NaN
1	4.0	5.0
2	NaN	NaN
3	8.0	9.0
4	3.0	4.0
5	7.0	7.0

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

Out[15]:
```
0    1.0
1    4.0
2    NaN
3    8.0
4    3.0
5    7.0
Name: a, dtype: float64
```

In [16]:
```
df.a.sum()
```

Out[16]:
```
23.0
```

In [17]:
```
df.cumsum()        # 累加
```

Out[17]:

	a	b
0	1.0	NaN
1	5.0	5.0
2	NaN	NaN
3	13.0	14.0
4	16.0	18.0
5	23.0	25.0

In [18]:
```
df.std()    # 标准差
```

Out[18]:
```
a    2.880972
b    2.217356
dtype: float64
```

In [19]:
```
df.var         # 方差
```

Out[19]:
```
<bound method DataFrame.var of      a    b
0  1.0  NaN
1  4.0  5.0
2  NaN  NaN
3  8.0  9.0
4  3.0  4.0
5  7.0  7.0>
```

In [20]:
```
df.max()
```

Out[20]:
```
a    8.0
b    9.0
dtype: float64
```

In [21]:
```
df.min()
```

Out[21]:
```
a    1.0
b    4.0
dtype: float64
```

In [22]:
```
df.quantile(0.75)
```

Out[22]:
```
a    7.0
b    7.5
Name: 0.75, dtype: float64
```

In [23]:
```
df.describe()
```

Out[23]:

	a	b
count	5.000000	4.000000
mean	4.600000	6.250000
std	2.880972	2.217356
min	1.000000	4.000000
25%	3.000000	4.750000
50%	4.000000	6.000000
75%	7.000000	7.500000
max	8.000000	9.000000