



DataFrame

☐ Thống kê

Method	Description
info()	Trả về thông tin của DataFrame như số mẫu, kiểu dữ liệu...
head()/tail()	Hiển thị danh sách các dòng đầu/cuối của Dataframe
describe()	Hiển thị thông tin thống kê chung của Dataframe
sum()	Trả về tổng các giá trị của các cột
mean()	Trả về giá trị số trung bình của các giá trị cột số
median()	Trả về giá trị ở vị trí giữa của các giá trị cột số
std()	Trả về độ lệch chuẩn của các giá trị cột số
min(), max()	Trả về giá trị nhỏ nhất, lớn nhất
sample()	Trả về các mẫu ngẫu nhiên từ Dataframe



DataFrame

☐ Thống kê

	toan	ly	hoa
0	3.5	4.0	5.0
1	8.0	9.0	9.0
2	7.0	8.5	9.5
3	4.0	5.5	8.0
4	8.5	6.0	7.0

```
df.mean(axis=1) # tính điểm trung bình theo học sinh
```

```
0    4.166667
1    8.666667
2    8.333333
3    5.833333
4    7.166667
dtype: float64
```

```
df.mean(axis=0) # tính điểm trung bình theo môn học
```

```
toan    6.2
ly      6.6
hoa     7.7
dtype: float64
```



DataFrame

❑ Thống kê

```
lst = [['Tom',25,1200,100], ['Mike',27,1500,100], ['Rose',32,3000,101], ['Alice',26,1200,101], ['Bill',25,1500,101]]
df = pd.DataFrame(lst, columns=['name','age','salary','depid'])
df
```

	name	age	salary	depid
0	Tom	25	1200	100
1	Mike	27	1500	100
2	Rose	32	3000	101
3	Alice	26	1200	101
4	Bill	25	1500	101

```
# thống kê chung
df.describe()
```

	age	salary	depid
count	5.000000	5.000000	5.000000
mean	27.000000	1680.000000	100.600000
std	2.915476	752.994024	0.547723
min	25.000000	1200.000000	100.000000
25%	25.000000	1200.000000	100.000000
50%	26.000000	1500.000000	101.000000
75%	27.000000	1500.000000	101.000000
max	32.000000	3000.000000	101.000000



DataFrame

□ Thống kê

```
# thống kê trên cột salary  
df['salary'].describe()
```

```
count      5.000000  
mean     1680.000000  
std       752.994024  
min      1200.000000  
25%      1200.000000  
50%      1500.000000  
75%      1500.000000  
max      3000.000000
```

```
df['depid'].value_counts()
```

```
101    3  
100    2  
Name: depid, dtype: int64
```

```
print('Thống kê theo salary')  
print('Count:', df.salary.count())  
print('Sum:', df.salary.sum())  
print('Mean:', round(df.salary.mean(),2))  
print('Median:', round(df.salary.median(),2))  
print('Min:', round(df.salary.min(),2))  
print('Max:', round(df.salary.max(),2))  
print('Variance:', df.salary.var())  
print('Standard Deviation:', df.salary.std())
```

Thống kê theo salary

Count: 5

Sum: 8400

Mean: 1680.0

Median: 1500.0

Min: 1200

Max: 3000

Variance: 567000.0

Standard Deviation: 752.994023880668



DataFrame

❑ Sắp xếp

```
# sắp tăng theo salary  
df.sort_values(by='salary')
```

	name	age	salary	depid
0	Tom	25	1200	100
3	Alice	26	1200	101
1	Mike	27	1500	100
4	Bill	25	1500	101
2	Rose	32	3000	101

```
# sắp giảm theo salary  
df.sort_values(by='salary', ascending=False)
```

	name	age	salary	depid
2	Rose	32	3000	101
1	Mike	27	1500	100
4	Bill	25	1500	101
0	Tom	25	1200	100
3	Alice	26	1200	101



DataFrame

❑ Sắp xếp

```
# sắp tăng theo depid và name  
df.sort_values(by=['depid', 'name'])
```

	name	age	salary	depid
1	Mike	27	1500	100
0	Tom	25	1200	100
3	Alice	26	1200	101
4	Bill	25	1500	101
2	Rose	32	3000	101

```
# sắp tăng theo depid và giảm theo salary  
df.sort_values(by=['depid', 'salary'], ascending=[True, False])
```

	name	age	salary	depid
1	Mike	27	1500	100
0	Tom	25	1200	100
2	Rose	32	3000	101
4	Bill	25	1500	101
3	Alice	26	1200	101