



DataFrame

Thuộc tính	Ý nghĩa
dtypes	Kiểu dữ liệu các cột của dataframe
columns	Danh sách các cột
axes	Danh sách các nhãn dòng và các tên cột
size	Số phần tử của dataframe
shape	Giá trị kiểu bộ (tuple) cho biết số phần tử của dòng, cột
index	Danh sách chỉ mục, nhãn dòng của dataframe
values	Mảng ndarray các giá trị của dataframe
loc, iloc	Cho phép truy xuất phần tử trong dataframe theo nhãn (loc) hay index (iloc)
...	



DataFrame

❑ Xem thông tin DataFrame

```
insurance_df = pd.read_csv('data/insurance.csv')
```

```
insurance_df.shape # có 1338 dòng, 7 cột
```

```
(1338, 7)
```

```
insurance_df.columns # in danh sách cột
```

```
Index(['age', 'sex', 'bmi', 'children', 'smoker', 'region', 'charges'], dtype='object')
```

```
insurance_df.dtypes # in danh sách cột và kiểu dữ liệu
```

```
age      int64
sex      object
bmi      float64
children int64
smoker   object
region   object
charges  float64
dtype: object
```



DataFrame

❑ Xem thông tin DataFrame

```
insurance_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1338 entries, 0 to 1337
```

```
Data columns (total 7 columns):
```

#	Column	Non-Null Count	Dtype
0	age	1338 non-null	int64
1	sex	1338 non-null	object
2	bmi	1338 non-null	float64
3	children	1338 non-null	int64
4	smoker	1338 non-null	object
5	region	1338 non-null	object
6	charges	1338 non-null	float64

```
dtypes: float64(2), int64(2), object(3)
```

```
memory usage: 73.3+ KB
```



DataFrame

❑ Xem thông tin DataFrame

```
insurance_df.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

```
insurance_df.tail()
```

	age	sex	bmi	children	smoker	region	charges
1333	50	male	30.97	3	no	northwest	10600.5483
1334	18	female	31.92	0	no	northeast	2205.9808
1335	18	female	36.85	0	no	southeast	1629.8335
1336	21	female	25.80	0	no	southwest	2007.9450
1337	61	female	29.07	0	yes	northwest	29141.3603



DataFrame

❑ Xem thông tin DataFrame

- Thông tin chung

```
insurance_df.describe() # các cột kiểu số
```

	age	bmi	children	charges
count	1338.000000	1338.000000	1338.000000	1338.000000
mean	39.207025	30.663397	1.094918	13270.422265
std	14.049960	6.098187	1.205493	12110.011237
min	18.000000	15.960000	0.000000	1121.873900
25%	27.000000	26.296250	0.000000	4740.287150
50%	39.000000	30.400000	1.000000	9382.033000
75%	51.000000	34.693750	2.000000	16639.912515
max	64.000000	53.130000	5.000000	63770.428010



DataFrame

❑ Xem thông tin DataFrame

- Thông tin chung

```
insurance_df.describe(include='object') # các cột kiểu object
```

	sex	smoker	region
count	1338	1338	1338
unique	2	2	4
top	male	no	southeast
freq	676	1064	364

