Pandas – gym example - statistics

Statistical functions can be used in DataFrame column series

```
In [27]: df['age'].mean()
Out[27]: 40.083333333333336

In [28]: df['height'].max()
Out[28]: 179

In [29]: df['weight'].min()
Out[29]: 42
```

```
In [30]: df.describe()
```

Out[30]:

	id	height	weight	age	hours	children
count	6.000000e+01	60.000000	60.00000	60.000000	60.000000	60.000000
mean	1.461910e+06	163.533333	59.95000	40.083333	3.600000	1.483333
std	2.932751e+05	9.307308	13.55144	10.469150	1.719282	1.214205
min	1.055891e+06	150.000000	42.00000	23.000000	1.000000	0.000000
25%	1.203819e+06	154.000000	50.00000	32.000000	2.000000	0.000000
50%	1.392896e+06	163.500000	56.00000	40.000000	3.000000	1.500000
75%	1.707049e+06	171.000000	65.25000	46.250000	5.000000	2.000000
max	1.998804e+06	179.000000	102.00000	59.000000	8.000000	5.000000

The DataFrame elements can be accessed in Python code

```
In [31]: def avg(col_name):
             s = 0
             n = len(df[col_name])
             for i in df[col_name]:
                 s = s + i
             avg = s / n
             return avg
```

```
In [32]: avg('age')
Out[32]: 40.083333333333336
```

Pandas – descriptive statistics

Function	Description
count	Number of non-NA observations
sum	Sum of values
mean	Mean of values
mad	Mean absolute deviation
median	Arithmetic median of values
min	Minimum
max	Maximum
mode	Mode
abs	Absolute Value
prod	Product of values
std	Bessel-corrected sample standard deviation
var	Unbiased variance
sem	Standard error of the mean
skew	Sample skewness (3rd moment)
kurt	Sample kurtosis (4th moment)
quantile	Sample quantile (value at %)
cumsum	Cumulative sum
cumprod	Cumulative product
cummax	Cumulative maximum
cummin	Cumulative minimum