DEFINITION
Correlation is the relationship between two variables. In other words, correlation means dependence or interdependence.

EXAMPLE
An example of a correlation may be the relationship between the amount of time devoted to learning and the grades obtained in the exams. We can assume that the more hours we study, the better grades we will get from exams.

Since correlation is an attribute, it does not have a formula. To calculate the correlation, we need to use one of the correlation coefficients. The Pearson correlation coefficient is the most commonly used.

FORMULA
$r_{XY} = cov(X,Y) / \sigma X \sigma Y$

r_{XY} - współczynnik korelacji Pearsona dla zmiennych X i Y / Pearson correlation coefficient for the X and Y variables
$cov(X,Y)$ - kowariancja zmiennych X i Y / covariance of the X and Y variables
$\sigma X \sigma Y$ - iloczyn odchyleń standardowych zmiennych X i Y / the product of the standard deviations of the variables X and Y

The correlation coefficient has no units, it always ranges from -1 to 1. Positive values mean positive correlation and negative values mean negative correlation. A value equal to 0 means there is no correlation (dependence) between the variables.

Positive correlation means that an increase of the values of one variable is accompanied by an increase of the values of the other variable, or a decrease of the values of one variable is accompanied by a decrease of the values of the other variable.

Negative correlation (anti-correlation) means that an increase in one variable is accompanied by a decrease in the other variable.

The closer the value of the correlation coefficient is to zero, the weaker the relationship between the variables. On the other hand, a value close to 1 or -1 indicates a strong correlation. For example, a value of -0.2 indicates a weak negative correlation.

To calculate the correlation coefficient, first calculate the covariance. It is similar to variance. The difference is that variance calculates the deviation of a single variable from the arithmetic mean, while covariance determines how two variables together differ from their means.

COVARIANCE: FORMULA
$covXY = \Sigma(x_i - x\_bar)(y_i - y\_bar) / n - 1$

Σ - sum of...
x_i - each value of the variable X
y_i - each value of the variable Y

x_bar - arithmetic mean of the variable X
y_bar - arithmetic mean of the variable Y
n - number of values of the X and Y variables <-- remember that the X and Y variables must have the same number of elements

COVARIANCE: UNITS OF MEASUREMENT
The units of the covariance are difficult to interpret. Since covariance examines two variables simultaneously, its units depend on the units of both variables.

In the example given earlier, we analyze two variables: test grades and time spent on learning (number of hours). So in this case, the units of covariance will be grade-hours.

Positive covariance values mean that two variables tend to move in the same direction. The values of both variables increase or decrease together.

Negative covariance values mean that two variables tend to move in inverse directions. One variable moves higher while the other decreases.

When the covariance is close to 0, it means that there is no linear relationship between the variables.

OTHER THINGS TO REMEMBER
Correlation does not imply causation.

Correlation is prone to outliers.