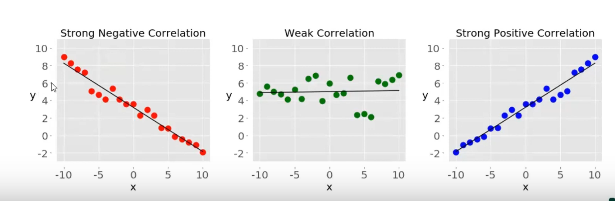
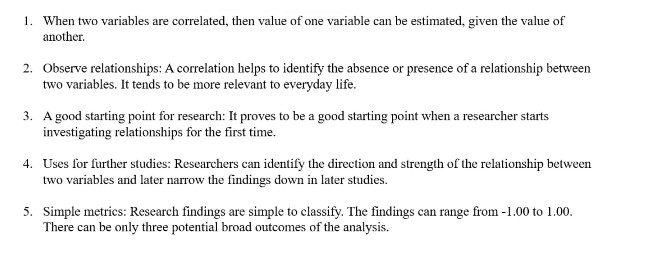
**Correlation:** In statistics, correlation or dependence is any statistical relationship, whether causal or not, between two random variables or bivariate data. In the broadest sense correlation is any statistical association, though it actually refers to the degree to which a pair of variables are linearly related.



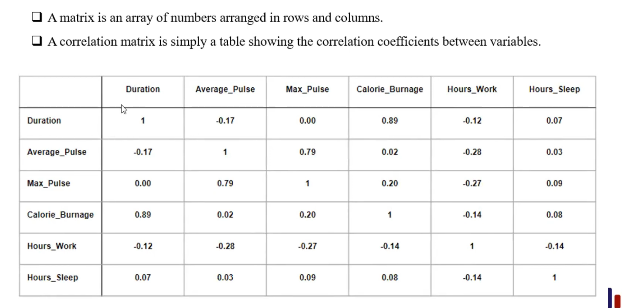
Once a correlation is known, it can be used to make predictions. When we know a score on one measure we can make a more accurate prediction of another measure that is highly related to it. The stronger the relationship between/among variables the more accurate the prediction.

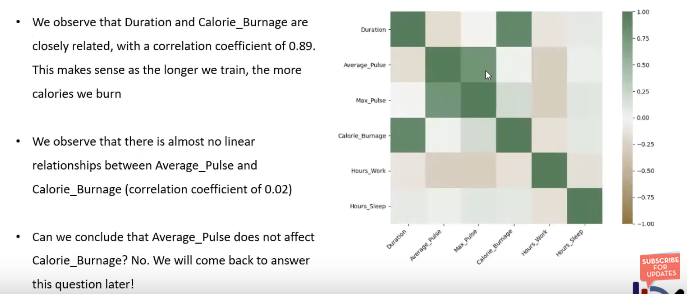


**Note:**

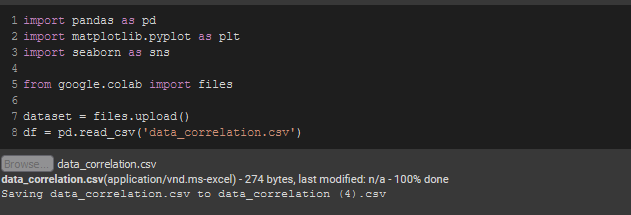
1. Correlation doesn’t cause caution. No axis has any affect on the other one.

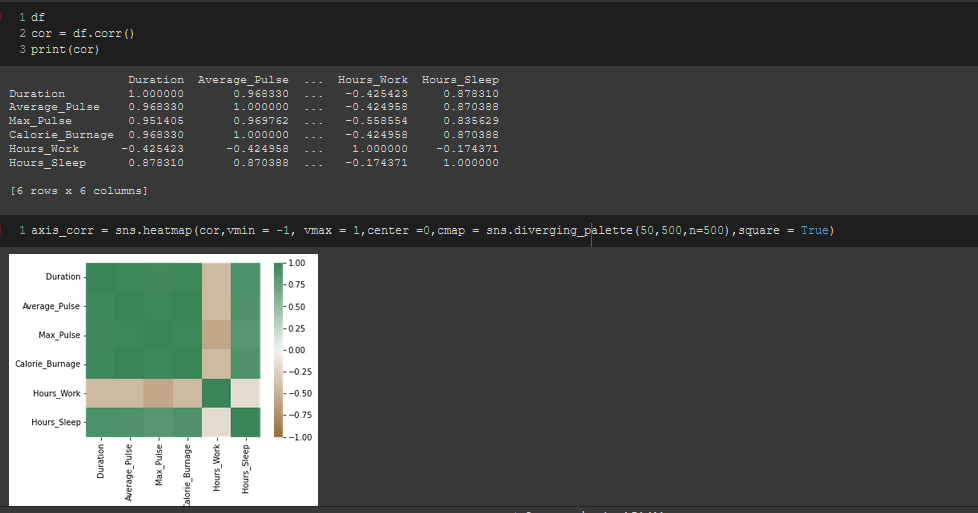
**Correlation matrix:** A Correlation matrix describes correlation among M variables. It is a square symmetrical MxM matrix with the (ij)th element equal to the correlation coefficient r\_ij between the (i)th and the (j)th variable. The diagonal elements (correlations of variables with themselves) are always equal to 1.00.





Dark colors represent strong correlation and weak colors represent weak correlation.



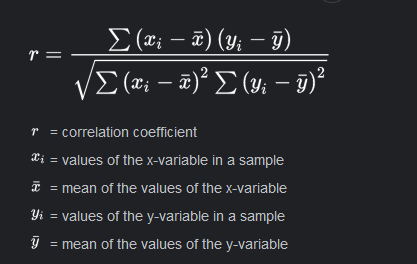


There are a lot of types of correlation available, 3 of the prominent types are,

1. **Pearson’s R:**

Pearson's r or Pearson product-moment correlation coefficient (PPMCC) is a measure of linear correlation between two sets of data. It is the ratio between the covariance of two variables and the product of their standard deviations; thus it is essentially a normalized measurement of the covariance, such that the result always has a value between −1 and 1.

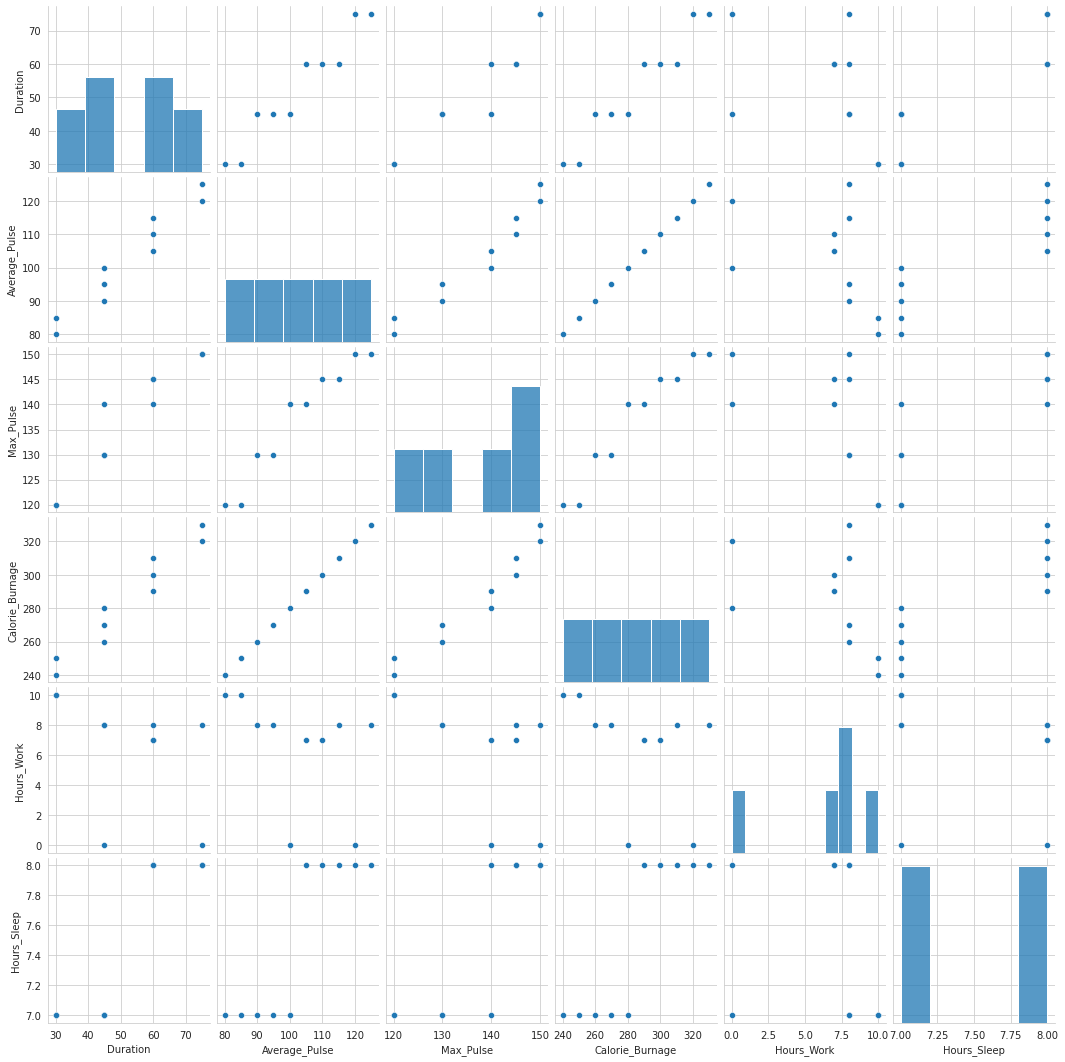
It can only identify if there’s any relation between 2 variables or not, It can’t identify how strong/weak that relationship is.



Pearson’s correlation aumes that all the datas are

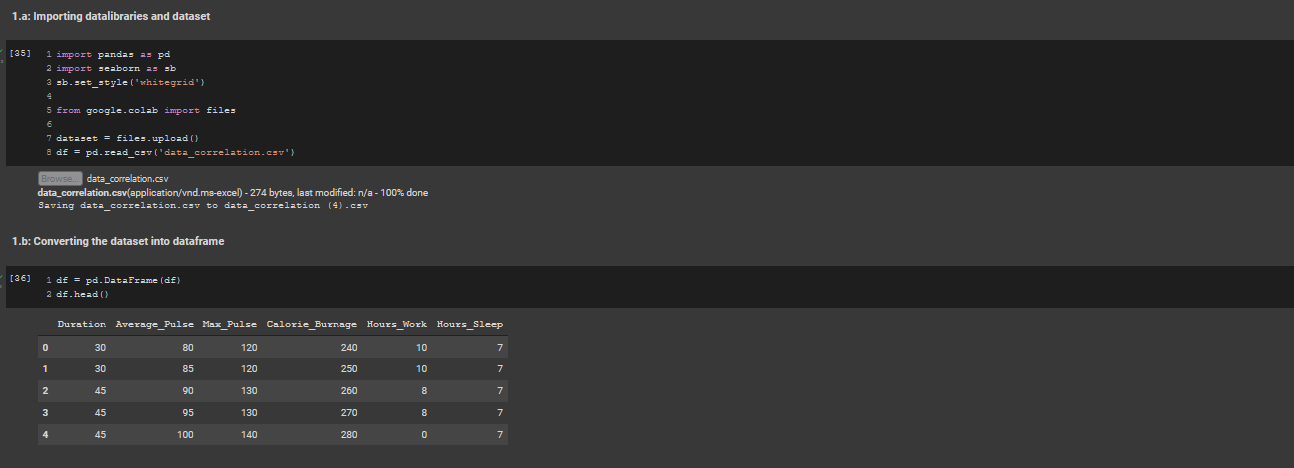
1. Normally distributed
2. Linearly related
3. Continuous and numeric

Data need to be normally distributed in order to generate a bell curve. The data below are not exactly normally distributed but close enough. Especially the diagonal wise figures.

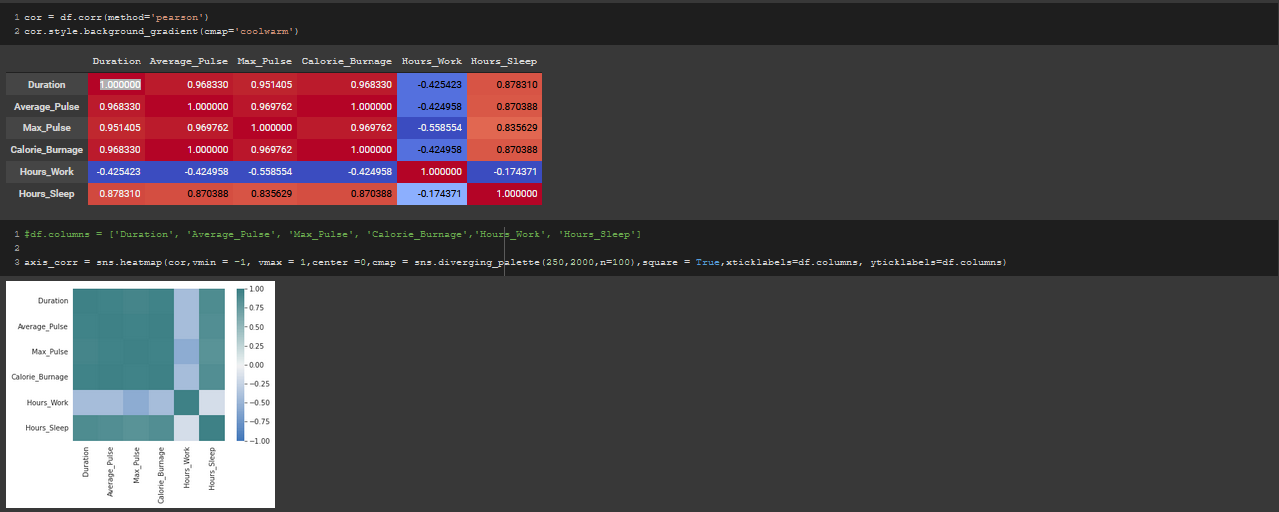


**Note:** These figures infer that the values are not continuous as there are gaps between them.

**Code:**



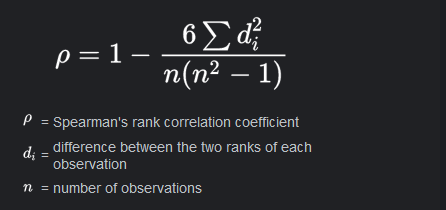
In pearson’s correlation, The output dataframe can be interpreted as for any cell, row variable correlation with the column variable is the value of the cell. As mentioned earlier, that the correlation of a variable with itself is 1. For that reason all the diagonal values are 1.00.



**Note:** Pearson’s correlation tells us if any correlation between 2 variables exist or not but not the strength of their correlation.

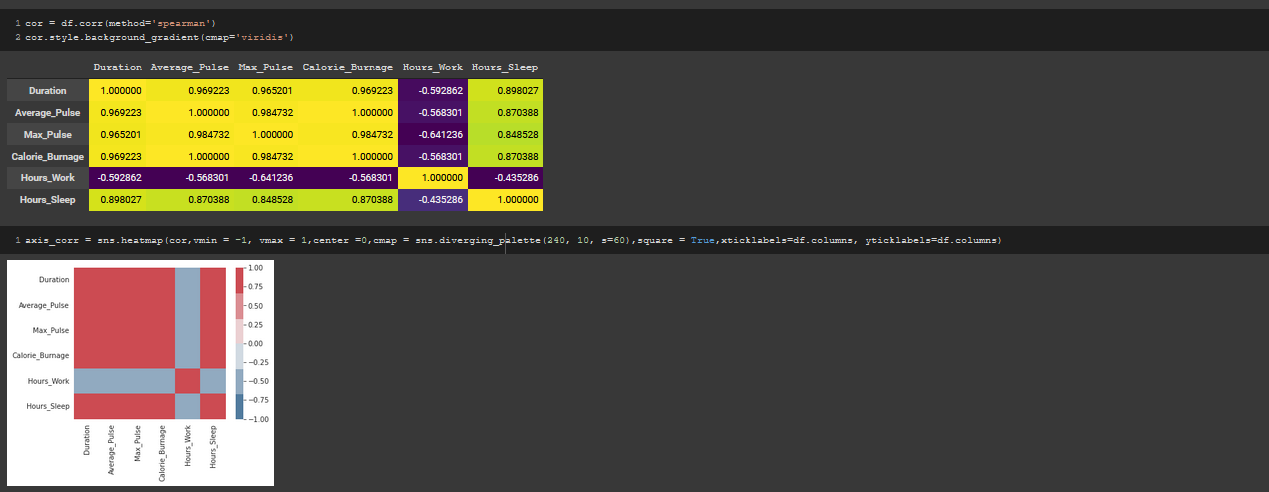
**2. Spearman’s Rho:**

The Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables; while Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not).



In addition, because Spearman’s measures the strength of a monotonic relationship, your data has to be monotonically related. Basically, this means that if one variable increases (or decreases), the other variable also increases (or decreases).

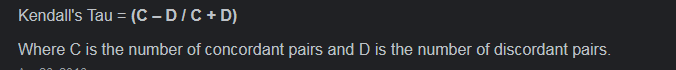
If there are no repeated data values, a perfect Spearman correlation of +1 or −1 occurs when each of the variables is a perfect monotone function of the other.

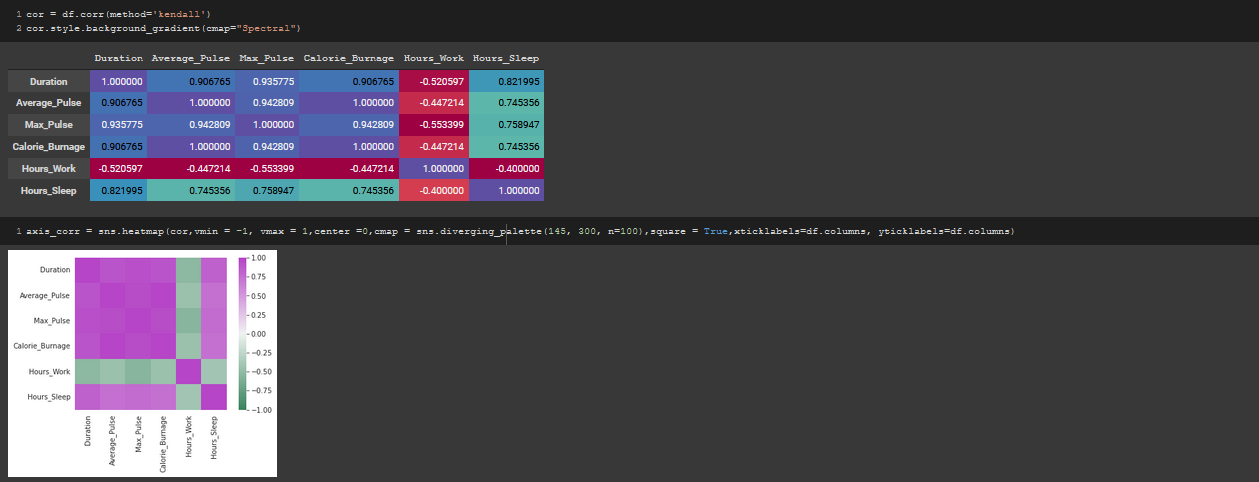


1.00 are perfect yellow, negative ones are purple and fractional values have 2 categories. If they are less than 0.9 or 0.94, They are lime and the one’s that are more than 0.94 are almost yello as in perfect 1 as 0.94 technically means 1.

**3. Kendall’s Tau:**

Kendall's Tau is a non-parametric measure of relationships between columns of ranked data. The Tau correlation coefficient returns a value of 0 to 1, where: 0 is no relationship, 1 is a perfect relationship





Kendall's rank correlation measures the strength and direction of association that exists (determines if there's a monotonic relationship) between two variables. Knowing this, testing for the presence of a monotonic relationship makes sense.