# Task 2 Python Documentation

Enrico Romano, Khan Gony Jany, Antonio Marco Vita

08/12/2021

## Dataset

After having created the dataset, we proceed with the analysis. First, we have removed some variables that are not particularly useful and went to check for missing values (NA).

## Data manipulation and data visualization

Then we moved on to data visualisation. The first thing we went to analyse was a correlation between elevation and temperature, to see if elevation was an indicator of temperature. The graph shows that elevation is not an indicator of temperature, but we construct a correlation matrix to check this assumption.

The correlation between temperature and elevation is close to 0, so our first hypotesis is confirmed. Next, we decided to make an analysis by countries, so we filtered the various countries (with the function df.loc()) on the basis of their latitude: Berlin, Paris, Rome, New York and Tokyo. The first analysis was to see which of the five countries has the lowest temperatures; as expected, London and Paris have the lowest temperatures and Tokyo and Rome the highest.

We then checked whether, in some countries, there are differences between predicted and apparent temperatures. Strong differences are shown in Berlin and New York. There are also some differences in Rome, although not as strong as in the above-mentioned cities. In conclusion, we went to check a possible correlation between humidity, elevation and precipitation. From the matrix, it can be seen that humidity does not seem to be correlated with precipitation, while it seems to be correlated with elevation.