LAB2 - Simple Linear Regression (continued)

In this lab, we will repeat our efforts to implement single linear regression on player age to experience (which was already done in LAB1), with small extra steps. This time, we will need two .csv files instead of one. First, we will have to *train* the algorithm using one of the datasets, i.e. we will use our simple linear regression algorithm on that dataset. Next, we will see how our estimation fares against the second dataset, we will *test* our regression line using that one. Two datasets to be used are given in the LAB2 folder.

Instructions:

- This lab is quite similar to the previous one, but instead of plotting the train dataset with the regression line, we will plot the *test* dataset (as a scatter plot) with the regression line. The regression line will be calculated using the *train* dataset.

- Use both .csv files, both as train and test datasets, which means there are going to be 2 different plots (in one run!). To be able to do this efficiently, we need to implement two different functions (as we did in LAB1), one which performs simple linear regression, and the other which plots the results.

- As a final extra step, calculate and display the residual sum of squares (RSS) for both cases. We are performing simple linear regression using the *least squares* method, which aims to minimize RSS. The calculation can be found here:

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

- The only outside packages available for us to use will be “pandas”, “numpy” and “matplotlib”.