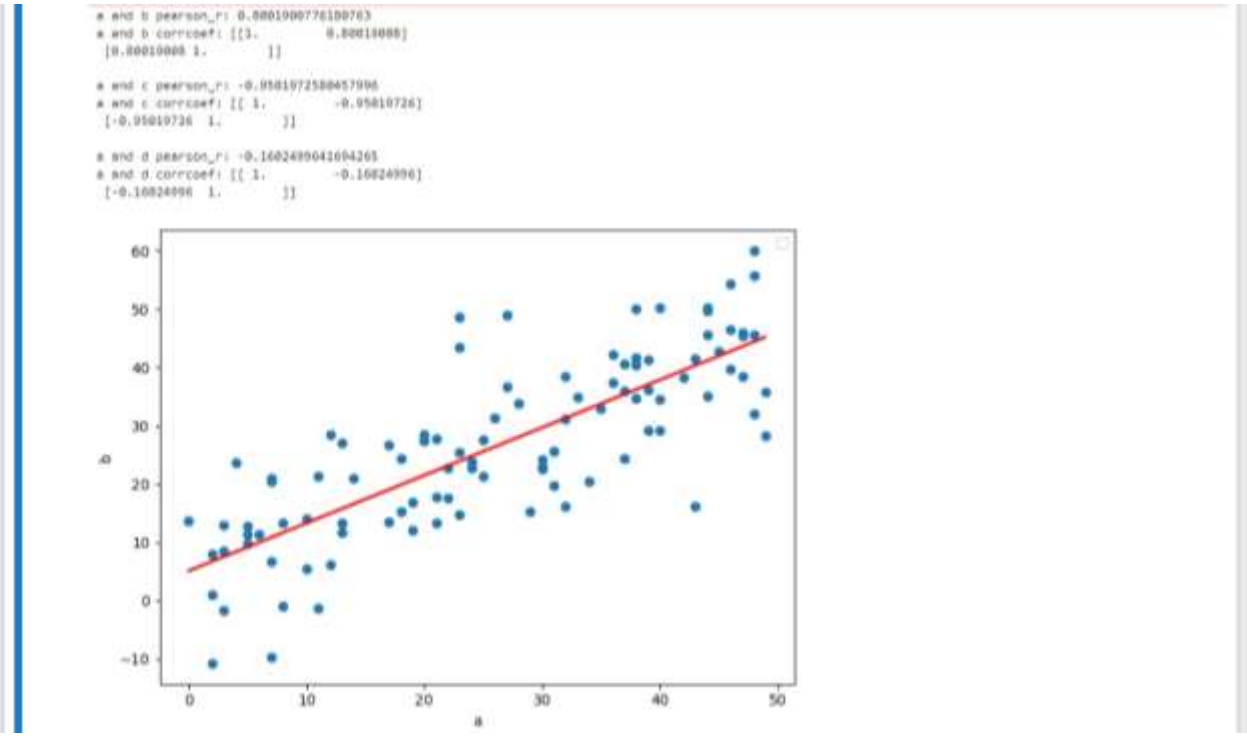
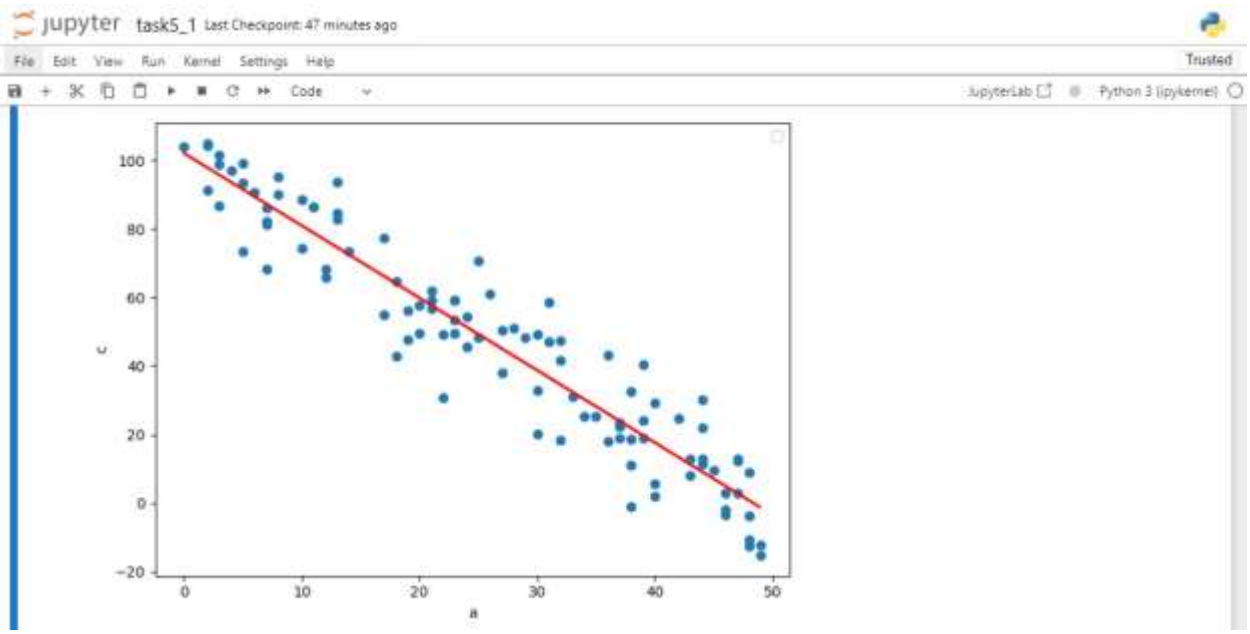
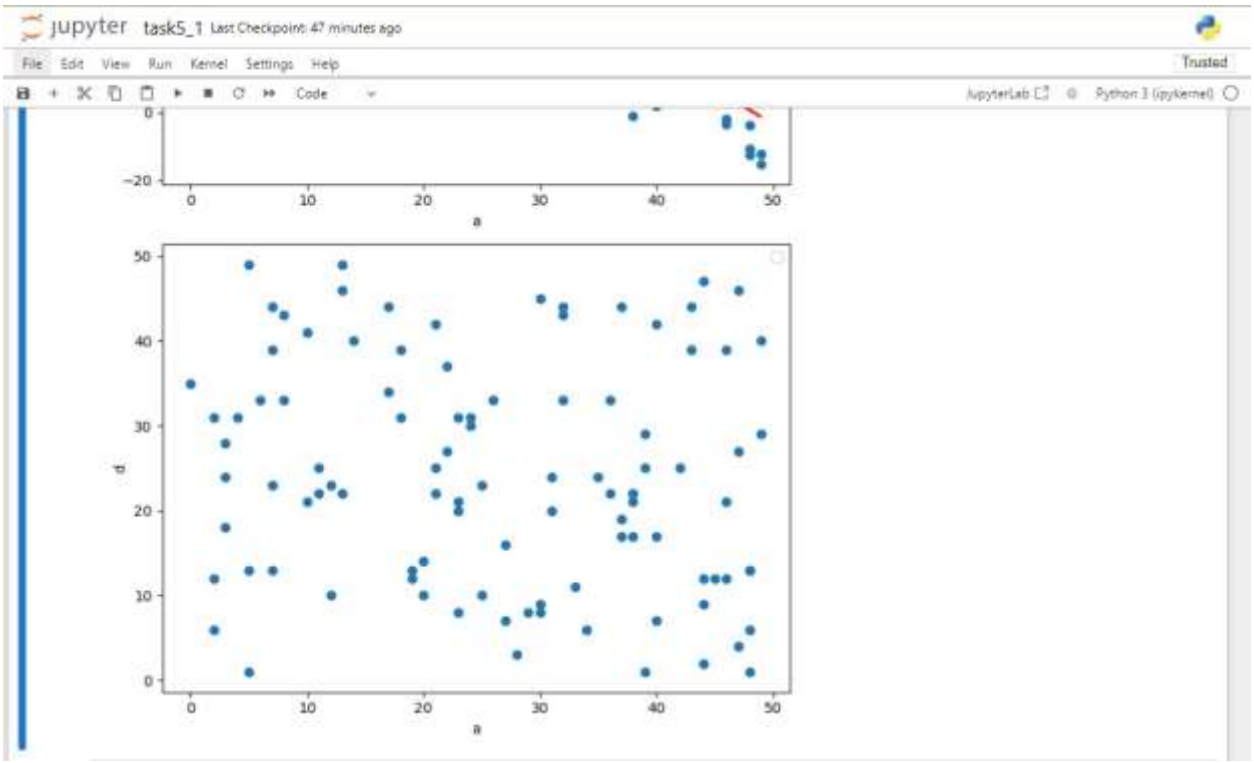
Pass Task 5.1P: Data correlation

Numpy implements a corrcoef() function that returns a matrix of correlations of x with x, x with y, y with x and y with y. We're interested in the values of correlation of x with y (so position (1, 0) or (0, 1)). The values of correlation of x with y might be Positive Correlation, Negative Correlation, or No/Weak Correlation.







The program examines the relationships among three variables ('b, 'c, and 'd') and the target variable ('a') in a CSV dataset. The correlation coefficient matrices and the results of the Pearson's correlation coefficient calculations are printed together.

The correlation matrix and Pearson correlation coefficient for each pair of variables ('a' and 'b', 'a' and 'c', 'a' and 'd') are calculated and printed using the print_correlation function.

Every pair of variables with 'a' on the x-axis receives a scatter plot from the plot_correlation function, which also, if the absolute correlation coefficient is higher than 0.5, may fit a line of best fit.

For every pair of variables, a subplot is made, and a line of best fit is optionally included in the scatter plots. Plt.tight_layout is used to provide appropriate spacing between the subplots, and plt.show is used to render the plots.