Supervised Learning - Simple Linear Regression:

1. A company manufactures an electronic device to be used in a very wide temperature range. The company knows that increased temperature shortens the life time of the device, and a study is therefore performed in which the lifetime is determined as a function of temperature. The following data is found:

Temperature in Celcius (t)	10	20	30	40	50	60	70	80	90
Life time in hours (y)	420	365	285	220	176	117	69	34	5

   a. Using Python, generates a scatterplot and investigates the correlation between the temperature value and the lifetime in hours.
   b. Find manually the linear regression line β0 and β1 values, provide the calculations you performed in your solution.
   c. Generate a line plot with Python to show the linear regression line you found
   d. Provide what will be the life time estimation in case the temperature will be 105, 45, 85

2. The height (sidewalk to roof) of notable tall buildings in America is compared to the number of stories of the building (beginning at street level).

Height (in feet)	Stories
1050	57
428	28
362	26
529	40
790	60
401	22
380	38
1454	110
1127	100
700	46

a. Using "stories" as the independent variable and "height" as the dependent variable, make a scatter plot of the data.
b. Does it appear from inspection that there is a relationship between the variables?
c. Calculate the least squares line. Put the equation in the form of: $\hat{y} = a + bx$
d. Find the correlation coefficient. Is it significant?
e. Find the estimated heights for 32 stories and for 94 stories.
f. Use the two points in (e) to plot the least squares line on your graph from (b).
g. Based on the above data, is there a linear relationship between the number of stories in tall buildings and the height of the buildings?
h. Are there any outliers in the above data? If so, which point(s)?
i. What is the estimated height of a building with 6 stories? Does the least squares line give an accurate estimate of height? Explain why or why not.
j. Based on the least squares line, adding an extra story adds about how many feet to a building?
k. What is the slope of the least squares (best-fit) line? Interpret the slope.

3. Use the **'Advertising.csv'** file and answer the following questions using Python:
    a. Investigate the correlation between each feature in the dataframe and the label. Generate separately scatterplot for each feature-label combination.
    b. For the next exercises use only the TV spends feature and the sales label.
    c. Generate a simple linear regression function on the TV spends vs Sales and find the beta coefficients of the linear regression line.
    d. Generate a line plot that shows the linear regression line you calculated in the previous exercise.
    e. Generate a simple scatterplot for TV spends vs Sales and add the linear regression line on top of the scatter plot.
    f. Generate a polynomial regression with 3 degrees and find all beta coefficient values of each feature.
    g. According to the beta coefficient values decide which features have an impact on the label and which features do not.