From the book I recommend: "Python for Data Analysis", please read the chapters 6. Data Loading, Storage, and File Formats and 9. Plotting and Visualization.

I also add useful links:

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.linregress.html?highlight=linear%20regression>

<https://matplotlib.org/3.1.1/api/_as_gen/matplotlib.pyplot.grid.html>

On this basis, exercises should be solved.

1. The supposition was made that the consumption of Coca-Cola is higher the higher the income of the population. To check the supposition, 11 families were drawn, for which the annual income per person in 2000 (variable X) and the annual drink consumption in liters per person (variable Y) were determined. The results are shown in the table. Check the guess by calculating covariance, correlation coefficient, simple regression coefficients. Make a graph showing the original data (scatter plot) and the fitted straight line. Describe the axes and give the title of the chart. Turn on the grid on the chart. Export the chart to a png file.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 350 | 420 | 200 | 800 | 1200 | 560 | 100 | 2000 | 1380 | 1000 | 340 |
| Y | 5 | 5,1 | 0 | 10 | 16 | 6,8 | 0 | 25 | 16 | 10 | 2 |

1. Using the mathematical pendulum model, students studied the value of the square of the vibration period from the length of the mathematical pendulum. Make a graph of the dependence of the pendulum length on the square of the vibration period, and based on the simplest squares fit method, determine the value of gravitational acceleration, calculate the value of the correlation coefficient. Describe the axes and give the title of the chart. Turn on the grid on the chart. Export the chart to a png file.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| L [m] | 1,00 | 1,10 | 1,20 | 1,30 | 1,40 |
| T2 [s2] | 4,02 | 4,37 | 4,84 | 5,26 | 5,62 |

1. The following results were obtained during measurements of a certain size: 13.2, 13.9, 13.65, 14.1, 12.3, 15.6, 17.2, 14.3,14.9, 10.9, 11 Calculate basic statistical parameters (mean value, standard deviation), make a histogram for 5 classes.
2. The Excel Korona\_Pl.xlsx file contains SARS Covid-19 virus development data. Import date column and increment values. Make a bar chart of the variability of the patient's growth in the following days, followed by a cumulative chart of changes in disease (red points connected by a blue line). Describe the axes, chart names, draw the grid. Export charts to a png file. What was the average incidence of Covid-19 in Poland?
3. The GR\_logging.xlsx file contains gamma-ray logging data in subsequent geological layers in a well. Make a box plot, where mean values, medians, standard deviations or quartiles and extreme values will be marked for separate lithostratigraphy. Mark each box for a given lithostratigraphy with a different color. Export the chart to a png file. Describe the axes and graph.