Introduction to Data Science and AI, Sofia, 2023

Assignment 1: Introduction to Data Science and Python

**Introduction**

The assignment is about implementing Python code that creates a scatter plot of GDP per capita vs life expectancy. The used data is provided by <https://ourworldindata.org/>.

**Running the program**

See the program’s “README.md” in the main program directory.

**Example graphic**

GDP per capita vs Life expectancy, all countries, the year 2018.

When generated by the program the graphic is interactive and allows for zooming in and out and selecting only sections of it. This should be the way to work with it as any screenshots like the one below would be hard if impossible to understand.

A graph with many names

Description automatically generated with medium confidence

**Questions in the assignment:**

Which countries have a life expectancy higher than one standard deviation above the mean?

*The program calculates this and displays a list of all the countries. Built-in functions of the Pandas dataframe are used to calculate this.*

Which countries have high life expectancy but have low GDP? Motivate how you have chosen to define “high” and “low”.

*This can easily be seen in the scatter plot. The countries in the upper left corner of the graphic are the ones that have above-average life expectancy despite below-average GDP per capita.*

Does every strong economy (normally indicated by GDP) have high life expectancy?

*Also depictable by the graphic. It clearly shows that countries with tangibly bigger GDP per capita (the ones above 40k) all have life expectancy far above average.*