Use of Python for Data Science & A.I. - Exercises

Question 1: Create a table (data frame) with 4 columns:

id: 1, 2, 3, 4, 5

age: 15, 18, 11, 17, 20

weight: 54, 68, 41, 59, 65

temp: 36.5, 37.1, 36.4, 37.2, 36.9

Question 2:

a. Copy the 2nd column from the table into a variable. What data type does this variable have?
b. Show the 3rd row off the table
c. Calculate the minimum and maximum age
d. Calculate the sum of all weights and share that by the number of weights
e. Divide all weights by 1000
f. Convert all temperatures to Fahrenheit

Question 3:

a. Save the table in a file
b. Read in the table in IntelliJ (Python) as a data frame. Change something in it and save again
c. Read the table from the file

Question 4:

a. Delete the "id" column
b. Add a column "length" with data: 160, 175, 162, 169, 179
c. Calculate the BMI of each person in the table
d. Add this BMI as a column to the table
e. Retain only the weights that are smaller or equal to 60 from the weight column.
f. In the age column, replace all ages greater or equal to 18 by 19
g. Find all temperatures of persons over 16 years of age

Question 5: Determine the separator and sign for comma numbers in the file "sportData.csv". Then read the file "sportData.csv" and put the result in the variable "sport". Finally, show all the columns of the table, along with their type.

a. How many data rows are there in this file?
b. Determine all possible values that appear in the column "Sport".
   Hint: you can do this by converting the column to a Category object and then extracting the categories from it.
c. The column "Sport" apparently contains the values "none". These values are not a useful value and we would like to read them as a missing value.

Python uses different values to indicate that a value is missing. The most common is NaN.
Read the file back in, but find an option to specify which strings should be recognized as missing value.
Check if it has succeeded by looking at all possible values of the column "Sport".

d. Which columns contain missing (NaN) values? Search the Internet to find this.
e. Which rows contain missing (NaN) values? Provide a list of all row numbers that contain at least 1 missing value.
Hint: Iterate through the rows and call the `row.isna().any()` method. This indicates whether there is a missing value in the row.
f. Delete all rows with missing values and put the result in a new table. This can be done with an existing method on the data frame. Find out how this is possible.
g. What is the maximum of column "age"?
h. What is the sum of all values of column "times a week" that are greater than 1?

Question 6:

a. Create a function "`replace(myList, to_look_for, replacement)`" that looks for a value in a list and replaces it with another.
b. Test the function by replacing the value 30 with 42 in the list [10, 20, 30, 30, 30].
c. Try if your function also works on a numpy array and a pandas Series