Tomasz Kisiel

24-11-2024

Python Project

Goal: to train my skills gained during Python course at Data Science studies.

Program version: Jupyter Notebook 7.2.2

Data library : SASHELP/FISH – has 159 observations of fish, columns are named: Species, Weight, Length 1, Length 2, Length 3, Height, Width

Task 0

Import a dataset fish.sas7bdat to pandas dataframe.

Task 1

Calculate average length. Set as a new column Legth. Delete other length columns.

Task 2

Calculate average weight for each type of fish. Sort them in decreasing order. Present only 50 the heaviest fish.

Task 3

Perform rounding of height and width with precision to decimals.

Task 4

Add a new column which has three first letters of the name of each fish.

Task 5

Add a new column which has a random normal number from 1 to 12 assigned to each observation.

Combine the column from task 4 and task 5 in order to create a new column Code in a form of “3SME”. Count how many observations will have “5BRE” code.

Task 6

Read a new data table from txt file which contains fish name and geographic name according to the following scheme.

|  |
| --- |
| Smelt, Pacific Ocean  Pike, Lake Erie  Perch, Lake Balkhash  Whitefish, Atlantic Ocean  Parkki, Mediterranean Sea  Roach, Caspian Sea  Bream, Volga River |

Task 7

Merge the new data table with existing fish file.

Task 8

Save the complete fish table with 159 observations and new columns into an excel file.