The three stages of reproducible workflow are as follows:

1. Data input/collection
2. Data processing
3. Data analysis

For the given table, my workflow is as mentioned below:

**1.Data input/collection**

Graphical user interface

Description automatically generated

I made the given table into a **raw*.csv*** file and then load that file in a python environment using ***pandas*** library.

Text, letter

Description automatically generated

**2. Data processing**

Here the predictors are converted into numeric data type. Since ***Frailty*** is the target variable, we do not need to convert its data type and let it remain as it is.

Graphical user interface, text, application

Description automatically generated with medium confidence

Graphical user interface, text

Description automatically generated

After processing, I save the processed dataset in processed.csv file under processed\_data folder.

processed.to\_csv('/content/drive/MyDrive/Assignment\_1/Part\_1/processed data/processed.csv',index=False)

**3. Data analysis**

Chart, waterfall chart

Description automatically generated

This visualization represents the relationship of ***height, weight*** and ***grip strength*** with ***age*** of each observation.

Graphical user interface, text, application

Description automatically generated

The folder structure looks like this after ***data analysis*** stage