Principles Of Data Science

Assignment – 1

1Q) ) Based on the following table, design the three stages of reproducible workflow, includes the work you  
can do and the folder structure in each stage (reference study case in chapter 3). (5 points)

Height (Inches) Weight (Pounds) Age Grip strength Frailty  
65.8 112 30 30 N  
71.5 136 19 31 N  
69.4 153 45 29 N  
68.2 142 22 28 Y  
67.8 144 29 24 Y  
68.7 123 50 26 N  
69.8 141 51 22 Y  
70.1 136 23 20 Y   
67.9 112 17 19 N  
66.8 120 39 31 N

Ans) From the above table we can say that a predictive model for frailty can be built based on the 4 attributes (Height, Weight, Age, Grip strength). If new data was provided with those four attributes the frailty can be predicted. The three stages of reproducible workflow can be given as

Stage 1: Consider every attribute and explain about them, including how are they measured and how it could affect the results, save all the explanation in a text file (Readme.txt). Also save the about table as a csv file (frailty\_data.csv).

Stage 2: There are no invalid, irrational, or missing values after visualization of data (scatter plot). Therefore, we can train a predictive model using raw data as input directly.

Stage 3: Now fit the data into multiple frailty detection models and analyze the results produced from those models.

These are the 3 steps of reproducible workflow for the given table.