**Principles of Data Science**

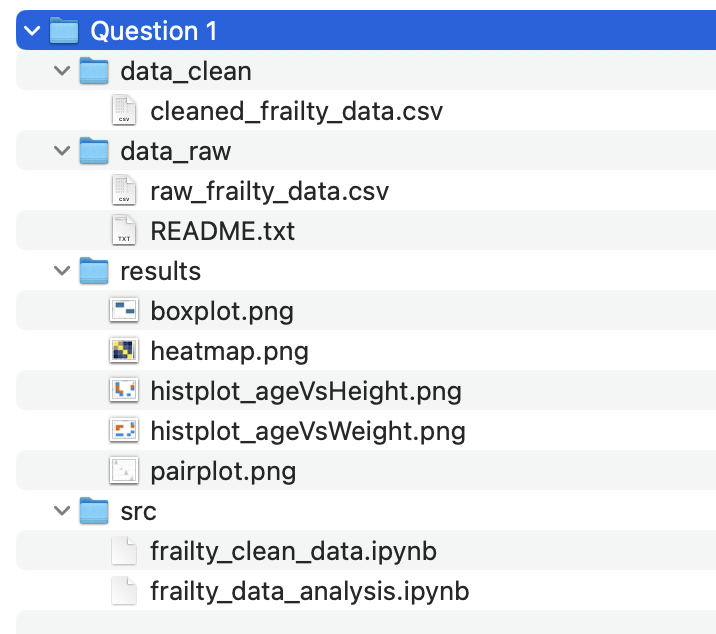
**Assignment-1**

Name: Sreevardhan Reddy Soma ID: 16352646

Question 1:

Step 1: Data Collection

* Given data set is stored in the form of CSV which makes analysis of the data easier as it will be in structured format.



* Here the data set is stored under data\_raw and README.txt file contains all the details about the dataset.

Given dataset is related to Frailty which is a physical weakness due to lack of health or strength. The dataset contains data of 10 female participants and five different attributes.

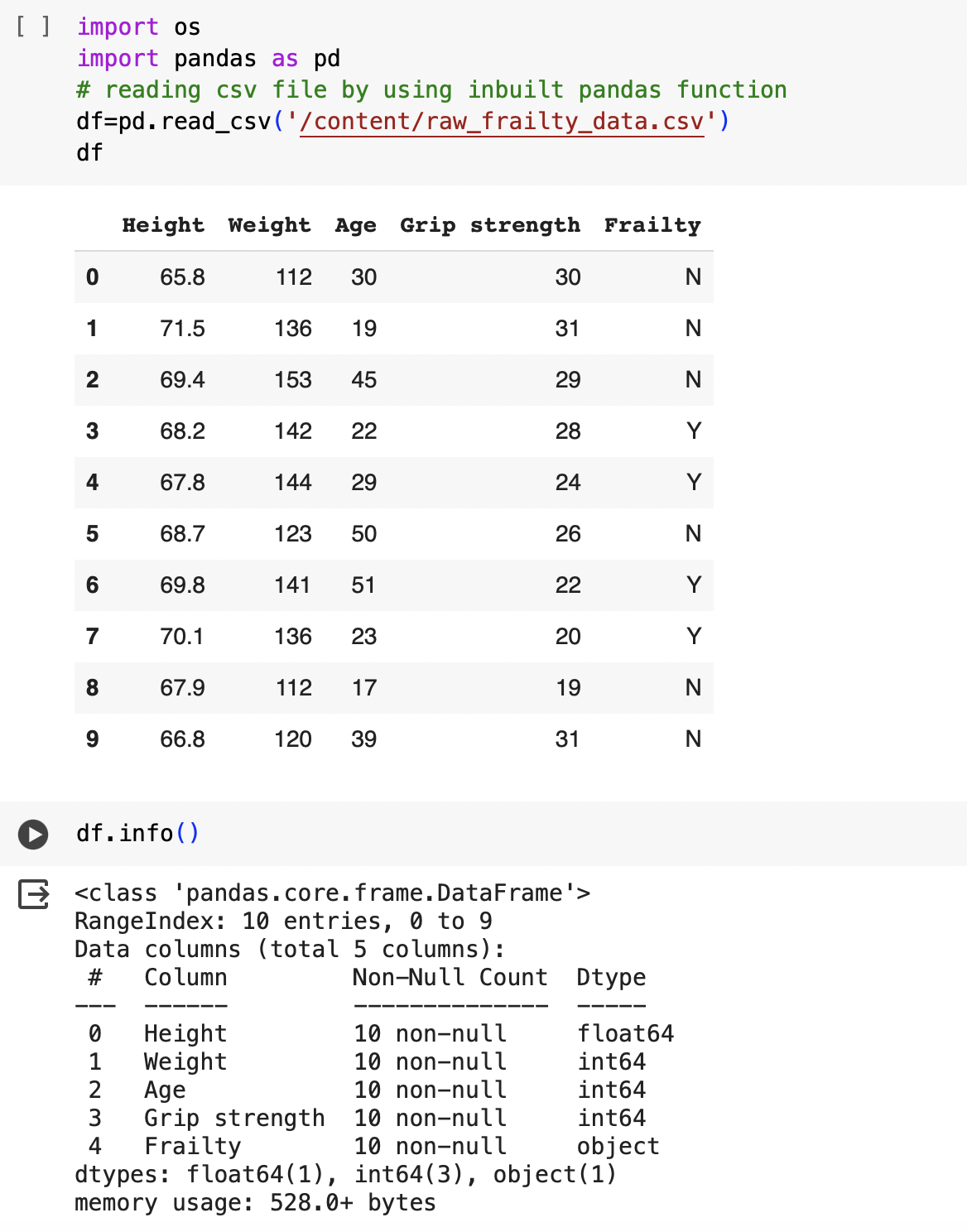
Description of dataset:

1. Height: Height of participants in inches
2. Weight: Weight of participants in pounds
3. Age: Age of participants in years
4. Grip Strength: The amount of static force that the hand can squeeze around a dynamometer which is in kilograms.
5. Frailty: indicates presence or absence of the frailty. Y for yes and N for no.

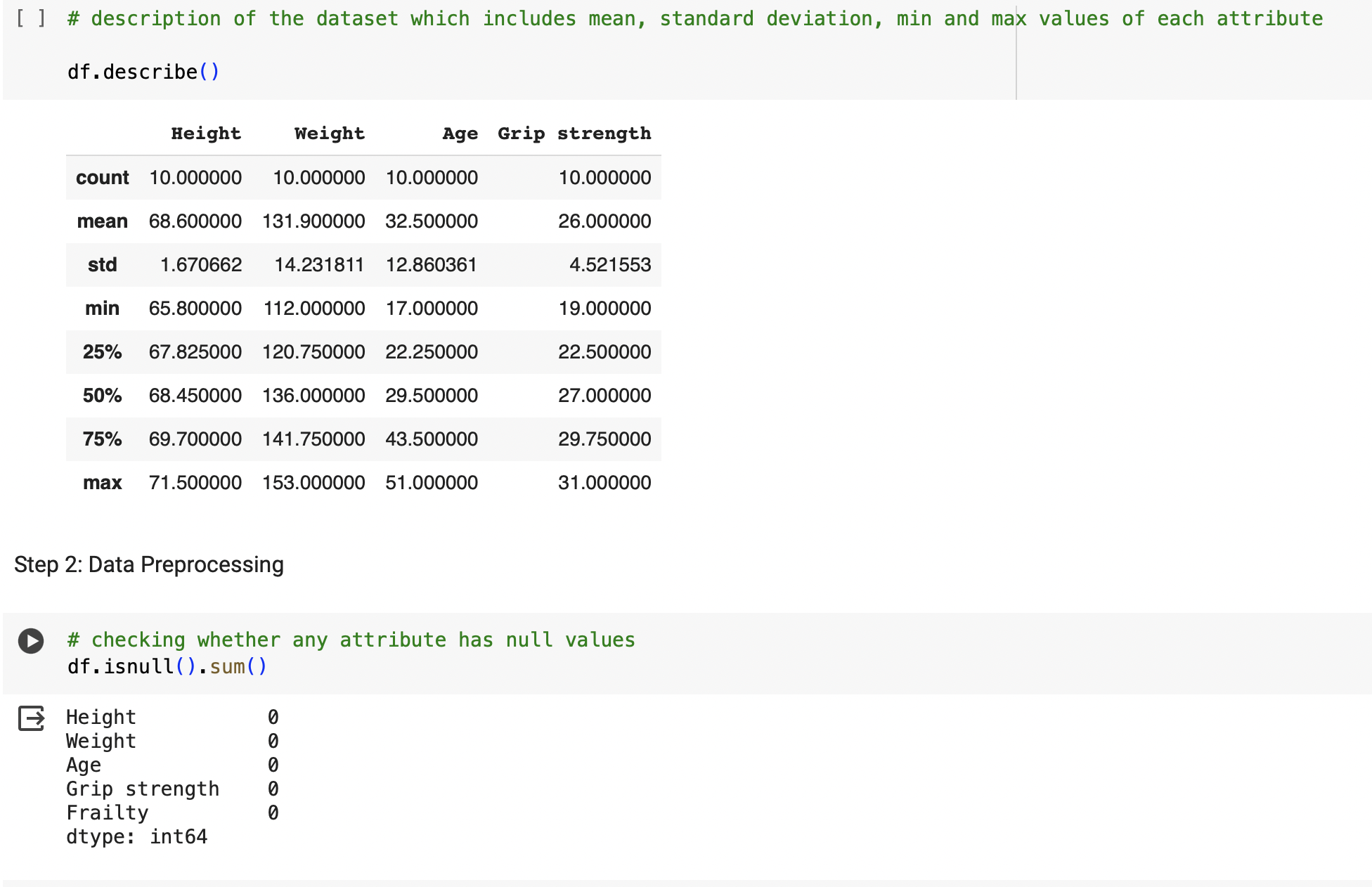
Data is provided in the Comma Separated Values(CSV) file.

Step 2: Data Processing

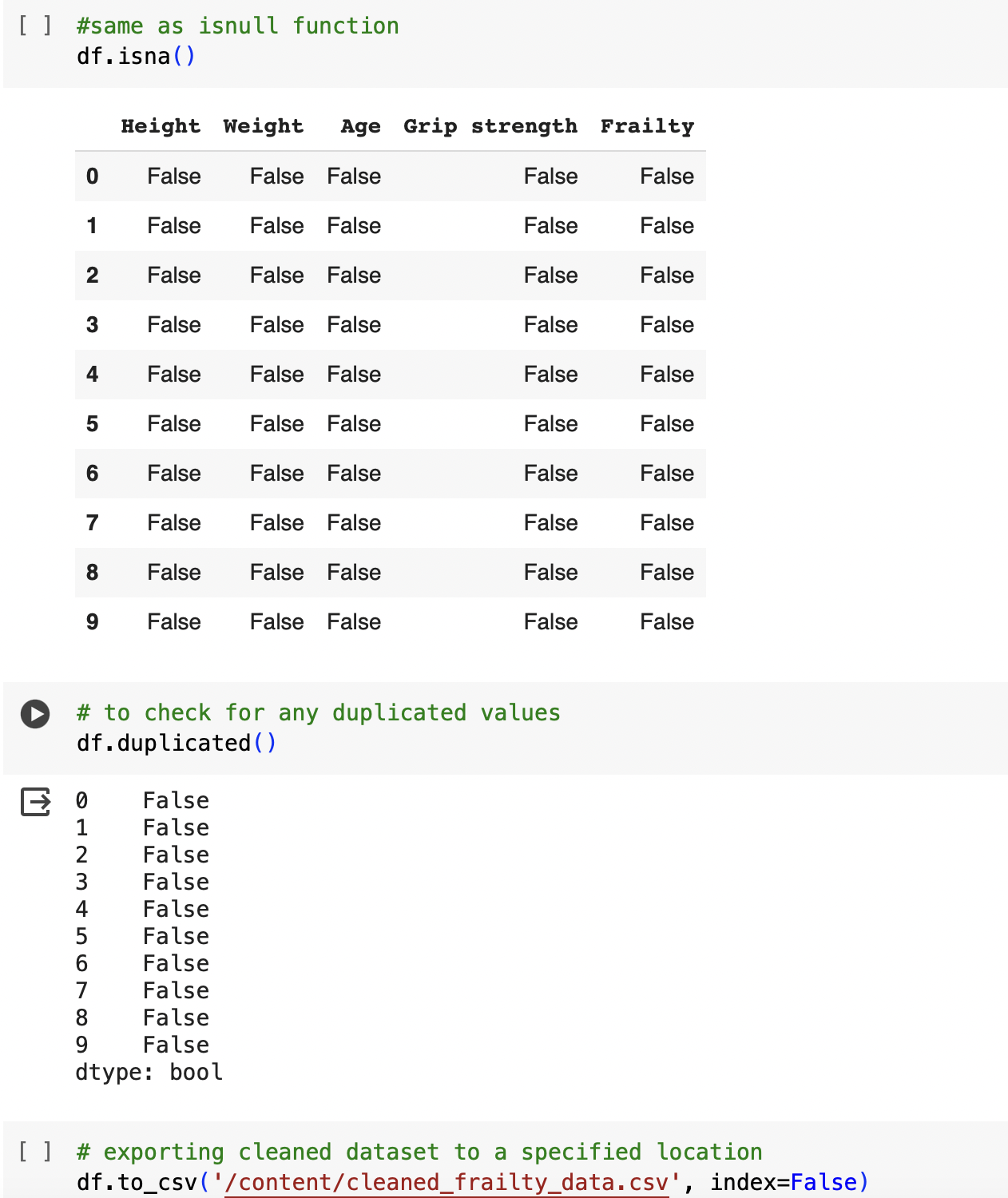
* Here we will perform some operations on the dataset to determine whether there were any null values or missing values present in our dataset. If so, we need to eliminate or replace them according to the problem statement.



* By using info(), we will get to know about the data type of the attributes and also if there were any null values present in the dataset. Here we can see that there were no null values.

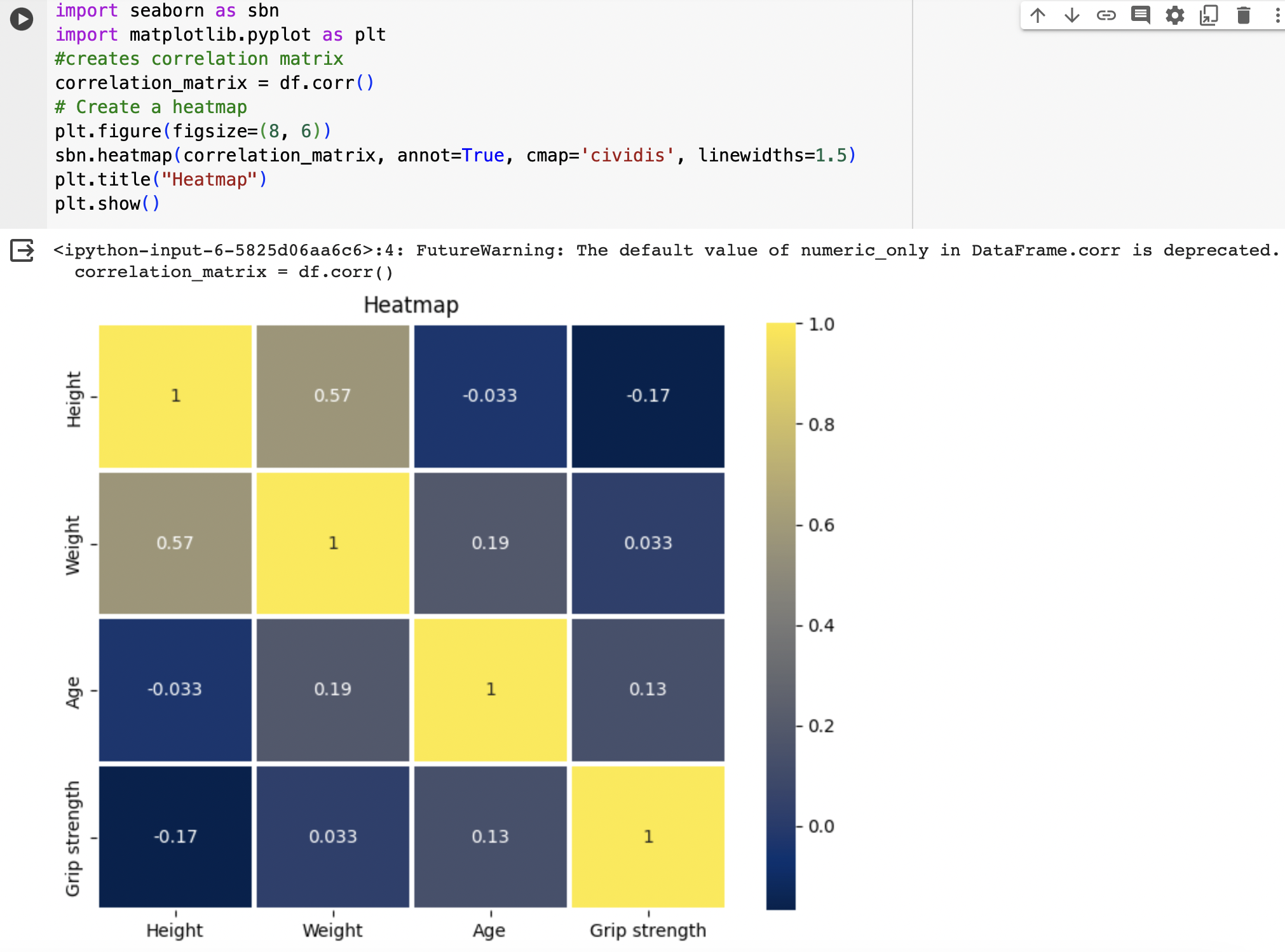


* describe() tells us the clear picture about the dataset that includes min and max values, standard deviation and also percentages at different stages which helps us in finding out the outliers.
* We can also check for null values using the isnull().sum().

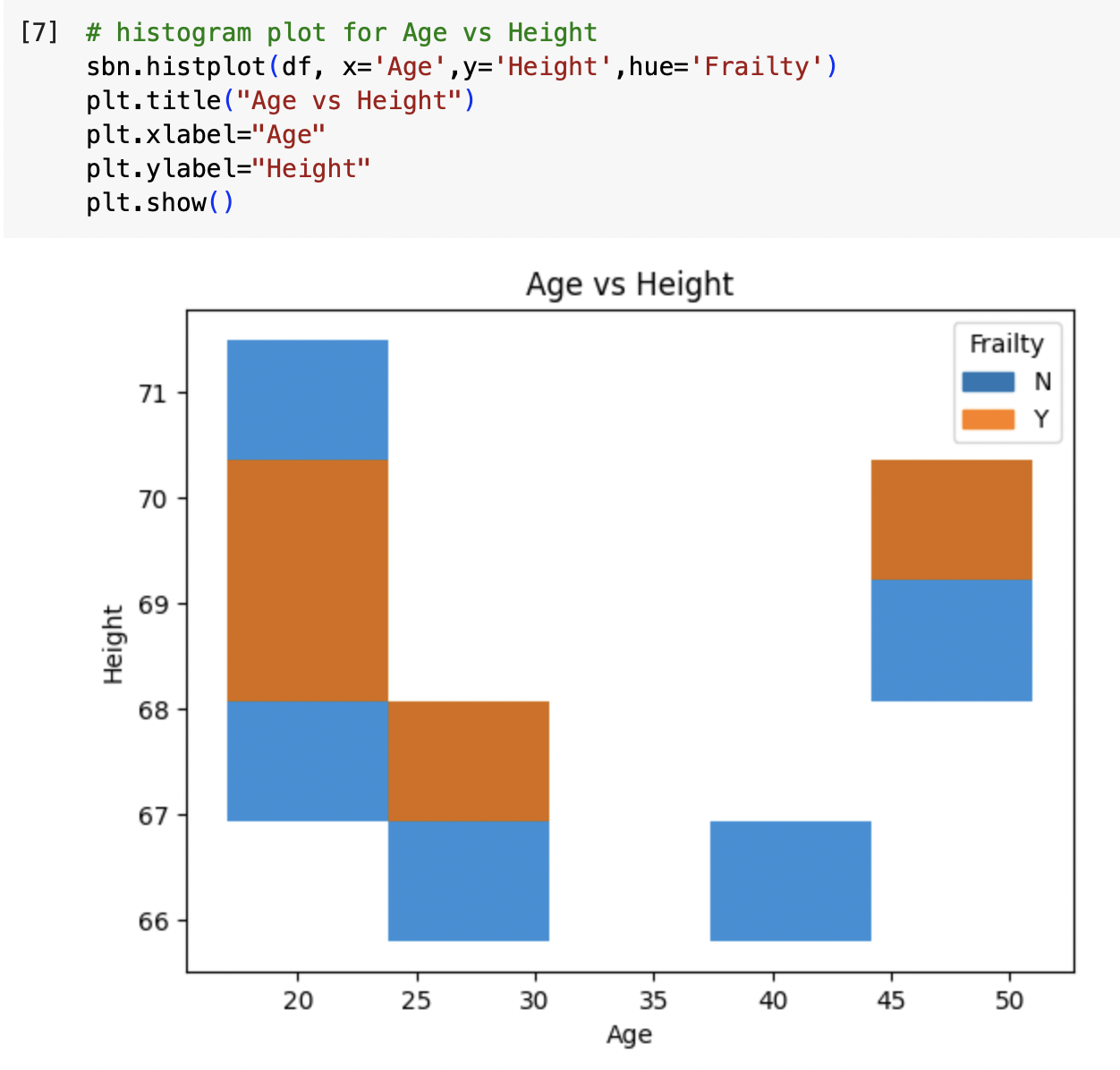


* isna() works the same as null function and gives us any null values present in the data set as true.
* After performing all these steps, we can download the cleaned dataset to appropriate location.

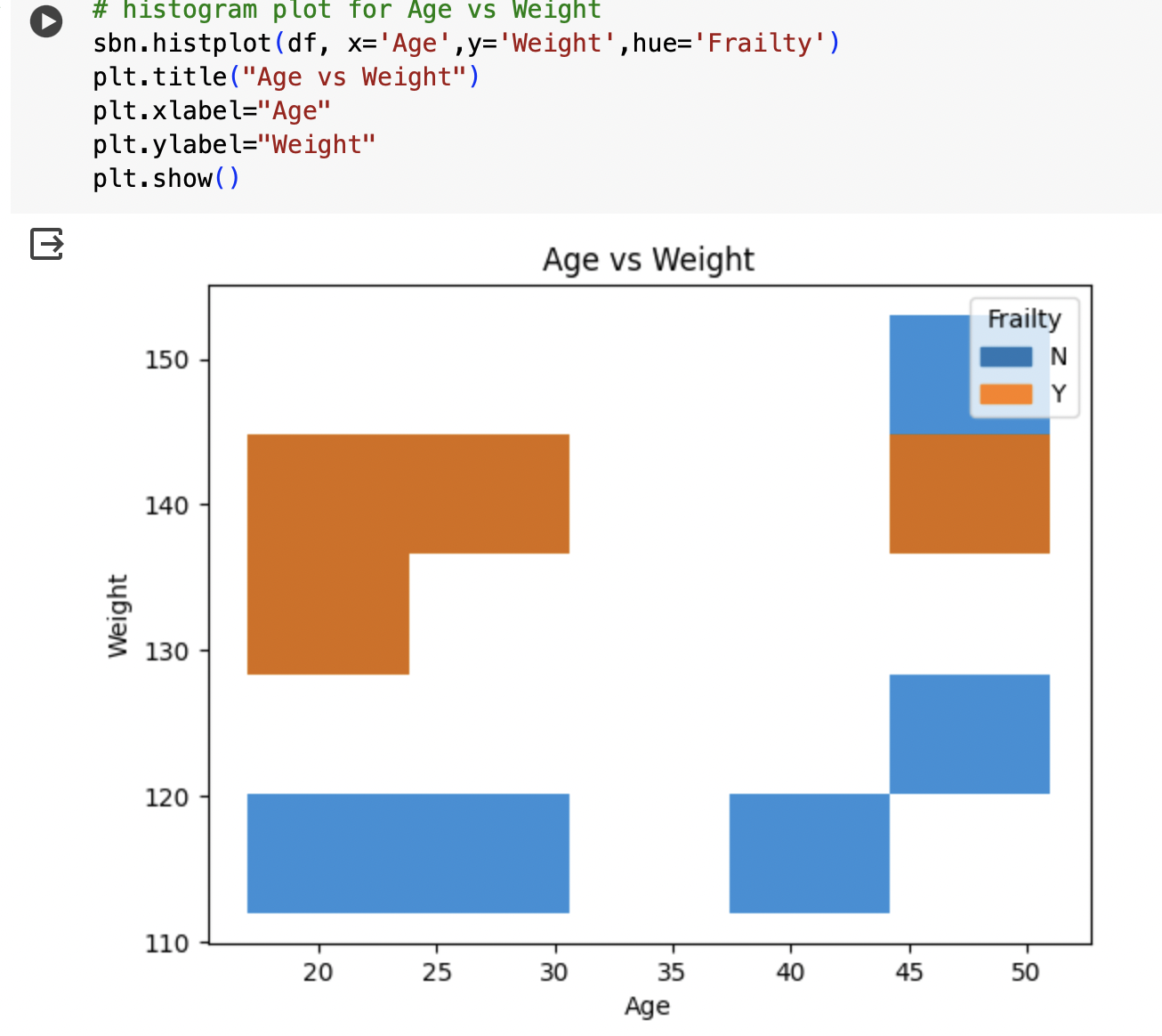
Step 3: Data Analysis



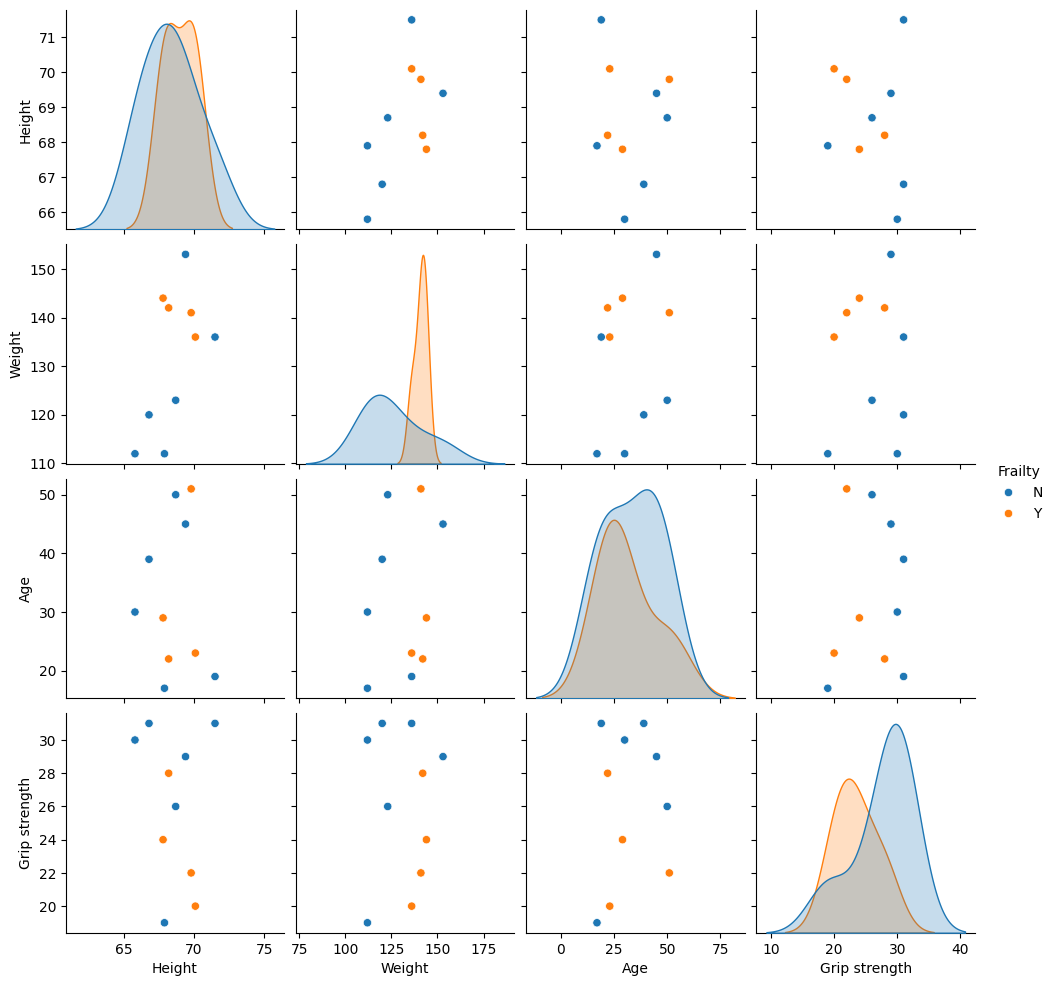
* The heat map provides us the relationship between the attributes. If the value in the heatmap is positive between two attributes, then we can say that those two attributes are positively correlated if not, negatively correlated. For example, the value between weight and height is 0.57 which is positive. That means, we can say that these two are positively correlated



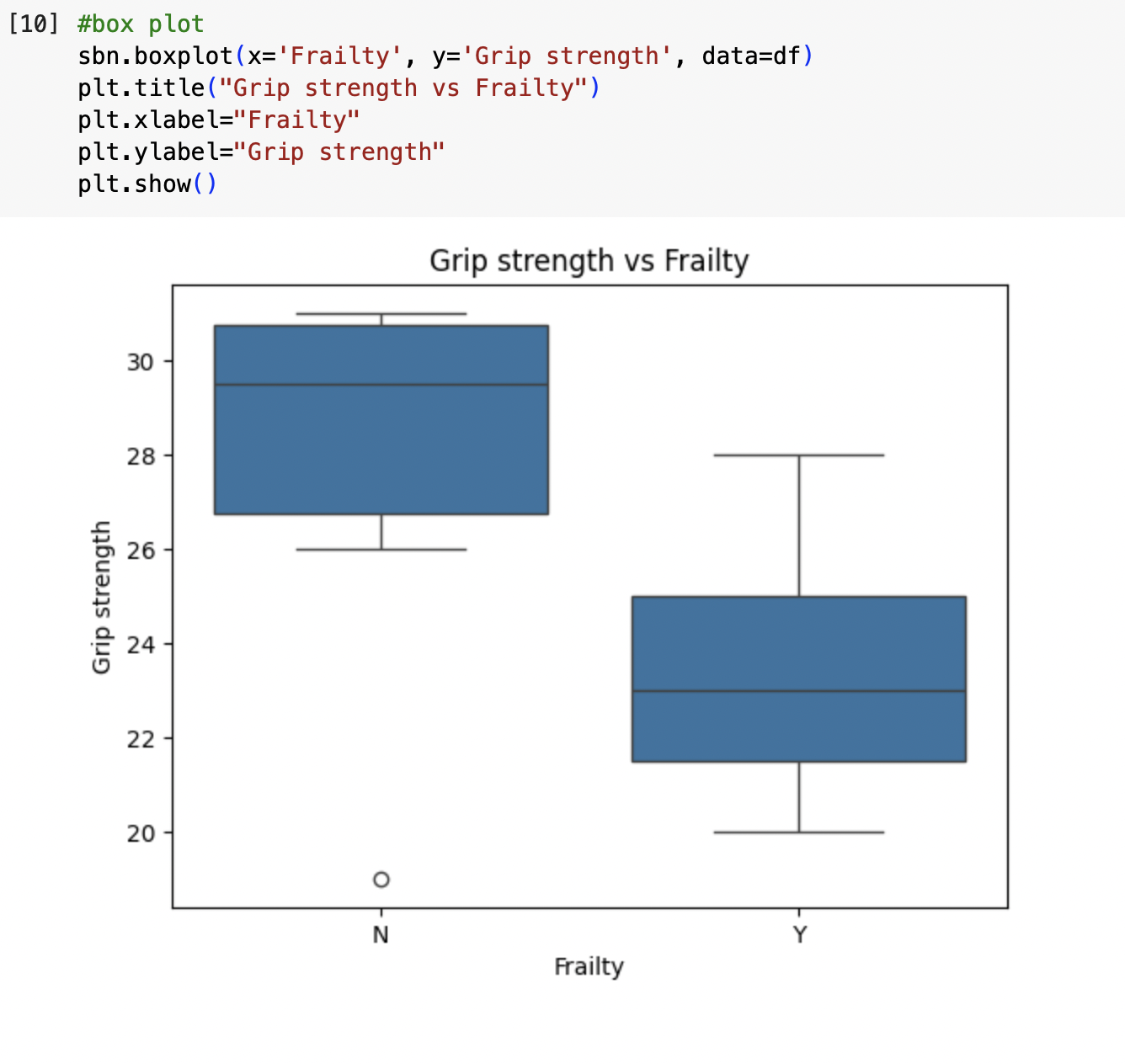
* From this analysis, we can infer that participants between height 68 to 70 at ages 20-25, they have high risk of getting Frailty. After the age of 25, the risk was observed for the participants of height range 67-68 and for the age group 45-50, risk was identified at the height range of 69-70. Since we have very little data, we can not conclude that this analysis satisfies with a larger dataset.



* Clearly, we can observe that people with a weight above 130 are more prone to the frailty at all ages than the people with weight below 130.



* This is a pairplot analysis which creates a scatter and dense plots between all the attributes. Participants with height 65-70, with weight more than 130, age 20-30 and with grip strength less than 26 are more prone to frailty.



* In this final analysis, a box plot between Grip Strength and Frailty is being shown.This tells us that the participants with grip strength between 21-25 are more vulnerable to the frailty.