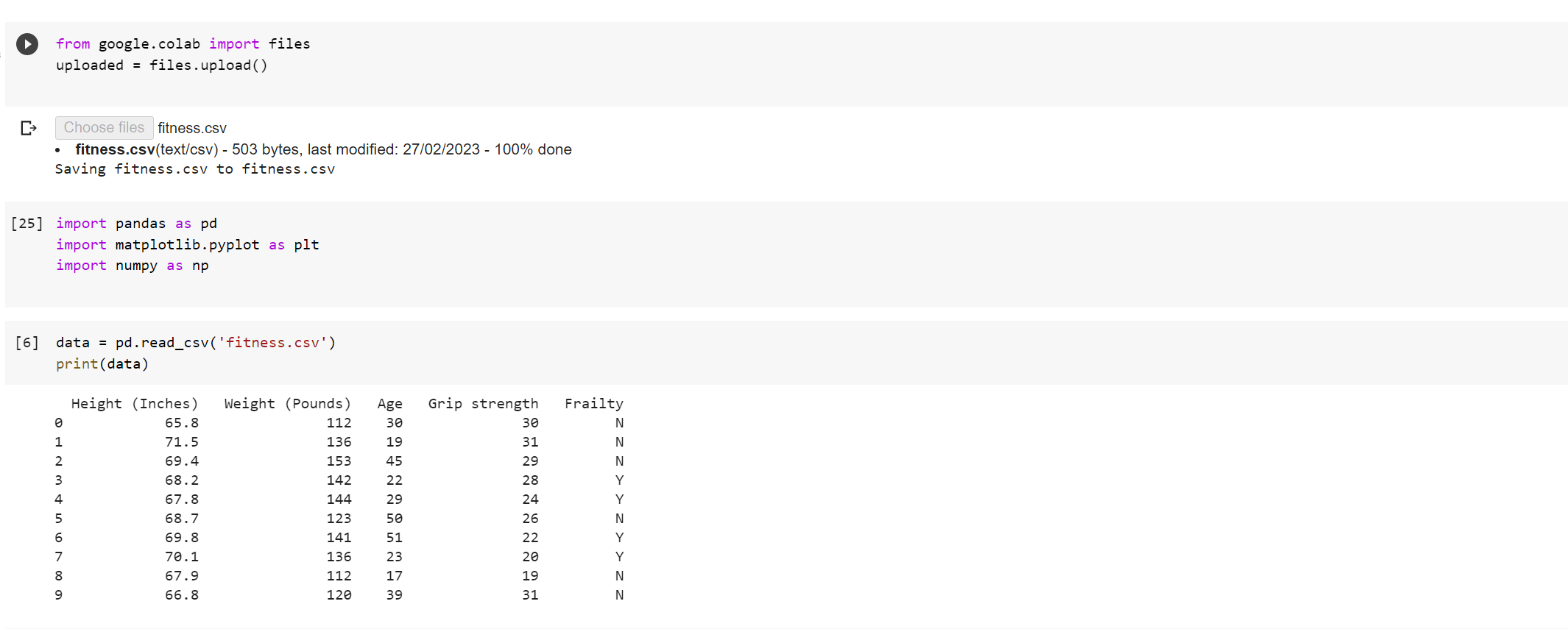
Assignment 1

1) 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)

A:

The three stages of a reproducible workflow are:

1. Data Acquisition: enter the data in the excel sheet and then save it in the format of .csv file.



2. Data preprocessing: In this stage, raw data is cleaned, organized, and transformed into a format suitable for analysis. This stage involves data cleaning, data validation, data normalization, data transformation, and data integration. The goal of this stage is to ensure that the data is reliable and reproducible.

3. Analysis: The goal of data analysis is to discover patterns, trends, and relationships within the data, and to make informed decisions or predictions based on the insights gained. 

We conclde that the people with age above 30 having strong( BMI ) i.e., 80% people are strong.And the people who is below 30 years are only 40% strong. So the people with age 30 and more is stronger than people with age below 30.

Source Code:

Upload CSV file

from google.colab import files

uploaded = files.upload()

import pandas as pd

import matplotlib.pyplot as plt

Reads and display the data from the file

data = pd.read\_csv('fitness.csv')

print(data)

Creates empty list and fill the colors with Green and Blue plots which differentiate two Frailty class

colors = []

for i in range(len(data)):

if str(data['Frailty\u202f\xa0'][i]) == 'N':

colors.append('Green')

else:

colors.append('Blue')

data.columns

Graph represents the scatter and visualization.

data.plot(x='Age\u202f\xa0',y='Grip strength\u202f\xa0', kind='scatter', color=colors)

plt.show()

data=data.sort\_values('Age\u202f\xa0')

print(data)