

# Principles of data science Assignment 1 - question 1

Hari chandra prasad Pasupuleti

16341883

1) Frailty is physical weakness; lack of health or strength. Reduced grip strength in females correlated

with higher frailty scores and vice versa. Hand grip strength can be quantified by measuring the amount

of static force that the hand can squeeze around a dynamometer. The force has most commonly been

measured in kilograms and pounds. The table below represents data from 10 female participants. The

Height is measured in inches, Weight in pounds, Age in years, Grip strength in kilograms. Frailty is

qualitative attribute indicated the presence or absence of the symptoms. 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 Weight 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

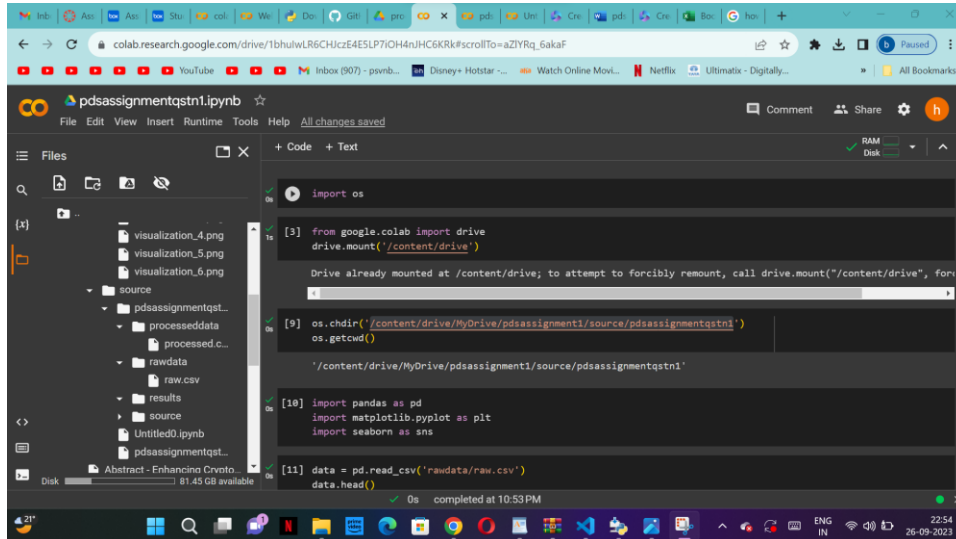
70.1 136 23 20 Y

67.9 112 17 19 N

66.8 120 39 31 N

Reproducible workflow consists of three steps they are collection of input data, processing of data, analysis of data.

## Step 1: Collection of input data.



```
import os

[3] from google.colab import drive
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

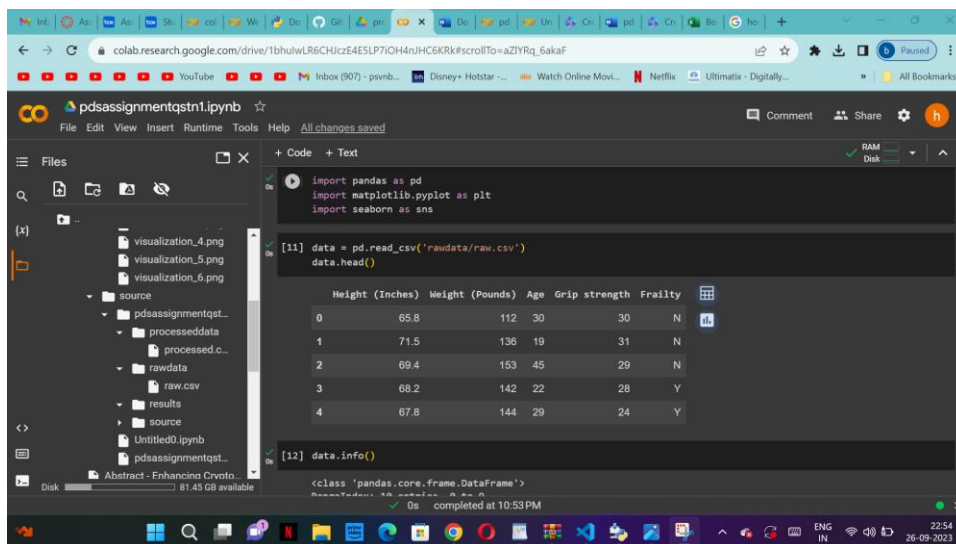
[9] os.chdir('/content/drive/MyDrive/pdsassignment1/source/pdsassignmenttstn1')
os.getcwd()

'/content/drive/MyDrive/pdsassignment1/source/pdsassignmenttstn1'

[10] import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

[11] data = pd.read_csv('rawdata/raw.csv')
data.head()
```

completed at 10:53 PM



```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

[11] data = pd.read_csv('rawdata/raw.csv')
data.head()
```

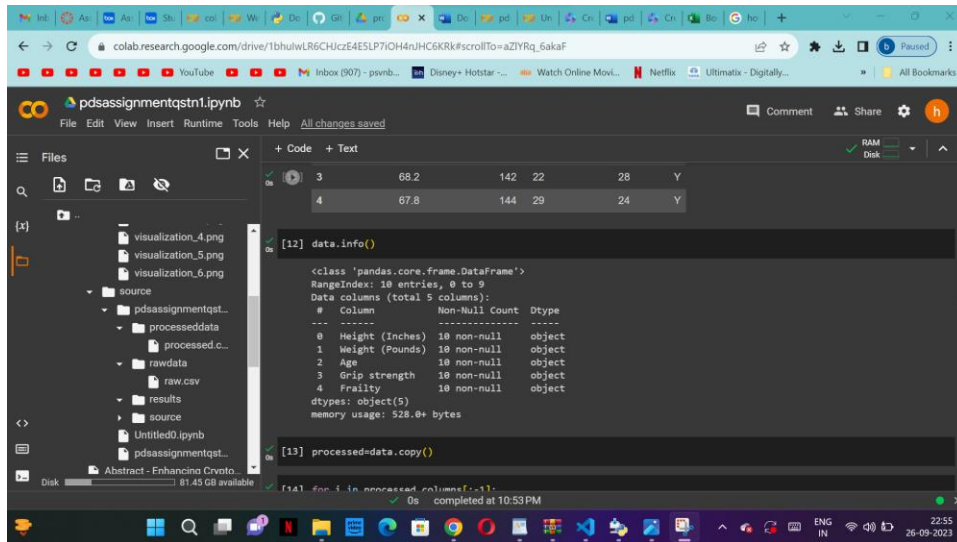
	Height (Inches)	Weight (Pounds)	Age	Grip strength	Frailty
0	65.8	112	30	30	N
1	71.5	136	19	31	N
2	69.4	153	45	29	N
3	68.2	142	22	28	Y
4	67.8	144	29	24	Y

```
[12] data.info()

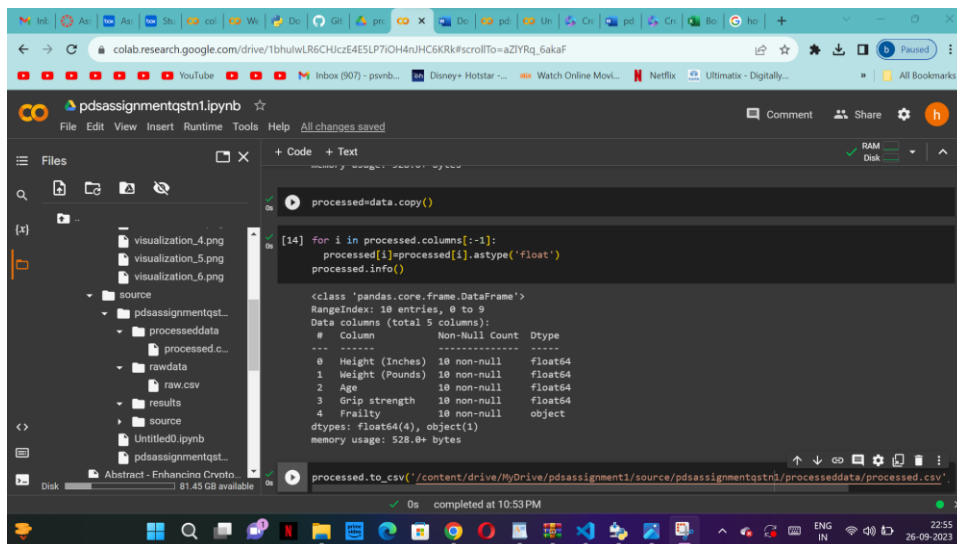
<class 'pandas.core.frame.DataFrame'>
```

completed at 10:53 PM

## Step 2: Processing of the data.

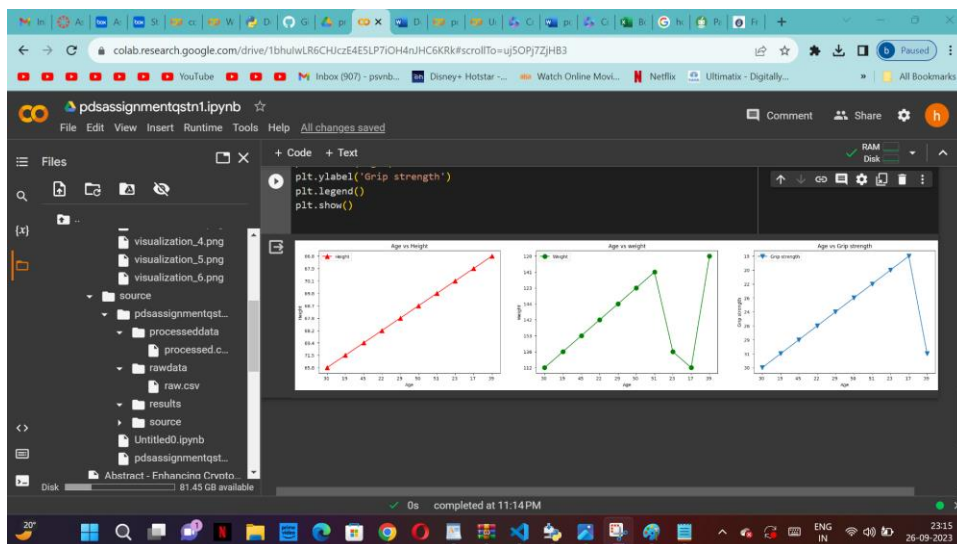
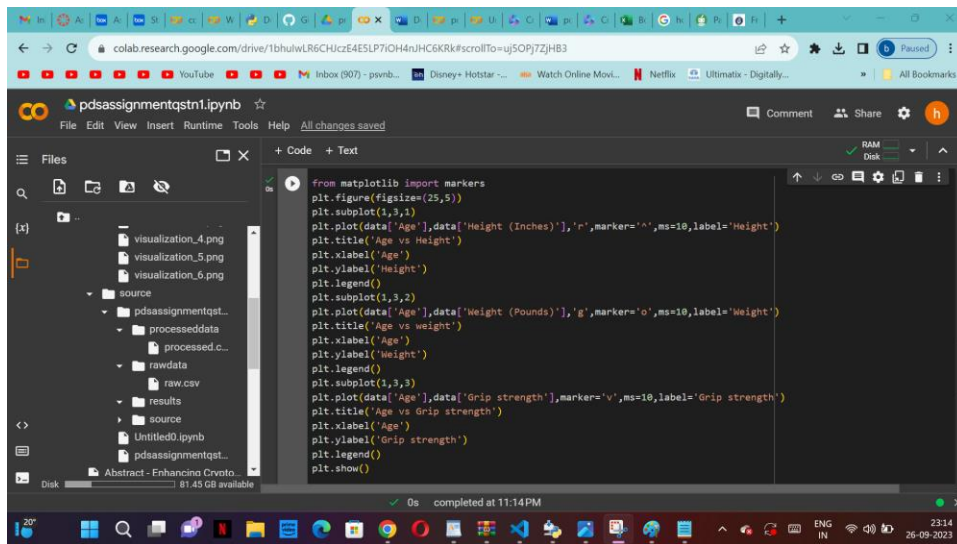


The predictors are changed to numeric data type here. Because Frailty is the goal variable, we don't need to transform it and can leave it as is.



Post processing the data it is saved as processed dataset in processed.csv file under processeddata folder.

Step 3: Analysis of data.



This graph depicts the association between each observation's height, weight, and grip strength and its age.

colab.research.google.com/drive/1bhulwLR6CHJczE4E5LP7OH4nJHC6KRK#scrollTo=aZiYRq\_6akaf

pdsassignmentqstn1.ipynb

Files

- visualization\_4.png
- visualization\_5.png
- visualization\_6.png
- source
  - pdsassignmentqstn1
  - processeddata
    - processed.c...
  - rawdata
    - raw.csv
  - results
  - source
  - Untitled0.ipynb
  - pdsassignmentqstn1

+ Code + Text

```
processed.to_csv('/content/drive/MyDrive/pdsassignment1/source/pdsassignmentqstn1/processeddata/processed.csv')

[ ] pdata = pd.read_csv('/content/drive/MyDrive/PDS_Assignment-1/Part-1/processed_data/processed.csv')

[ ] plt.figure(figsize=(20,5))
plt.subplot(1,4,1)
plt.boxplot(pdata['Height (Inches)'])
plt.xlabel('Height')
plt.subplot(1,4,2)
plt.boxplot(pdata['Weight (Pounds)'])
plt.xlabel('Weight')
plt.subplot(1,4,3)
plt.boxplot(pdata['Age'])
plt.xlabel('Age')
plt.subplot(1,4,4)
plt.boxplot(pdata['Grip strength'])
plt.xlabel('Grip strength')
plt.show()
```

0s completed at 10:53 PM

colab.research.google.com/drive/1bhulwLR6CHJczE4E5LP7OH4nJHC6KRK#scrollTo=aZiYRq\_6akaf

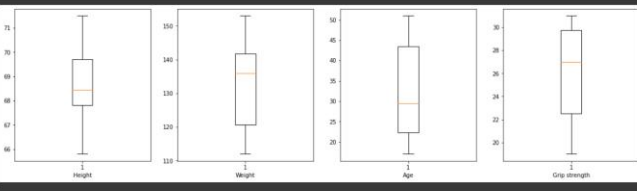
pdsassignmentqstn1.ipynb

Files

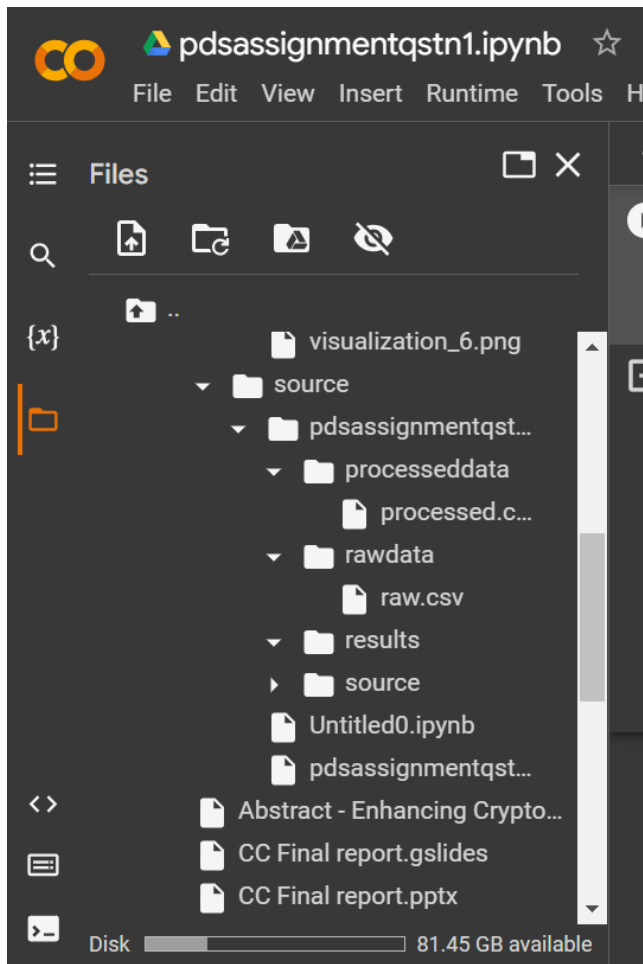
- visualization\_4.png
- visualization\_5.png
- visualization\_6.png
- source
  - pdsassignmentqstn1
  - processeddata
    - processed.c...
  - rawdata
    - raw.csv
  - results
  - source
  - Untitled0.ipynb
  - pdsassignmentqstn1

+ Code + Text

```
plt.boxplot(pdata['Age'])
plt.xlabel('Age')
plt.subplot(1,4,4)
plt.boxplot(pdata['Grip strength'])
plt.xlabel('Grip strength')
plt.show()
```



0s completed at 10:53 PM



The final file structure consisting of processed data, raw data, source file.