

Assignment 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

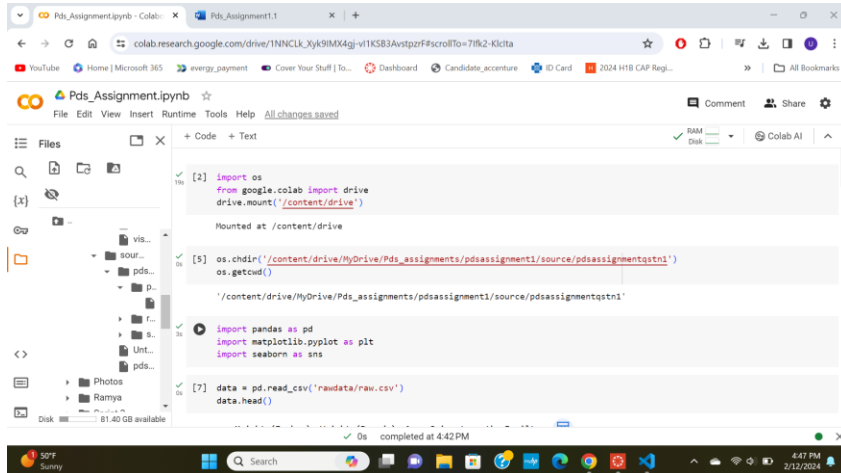
69.8 141 51 22 Y

70.1 136 23 20 Y

67.9 112 17 19 N

66.8 120 39 31 N

Step 1: Collection of input data



```
[2] import os
from google.colab import drive
drive.mount('/content/drive')

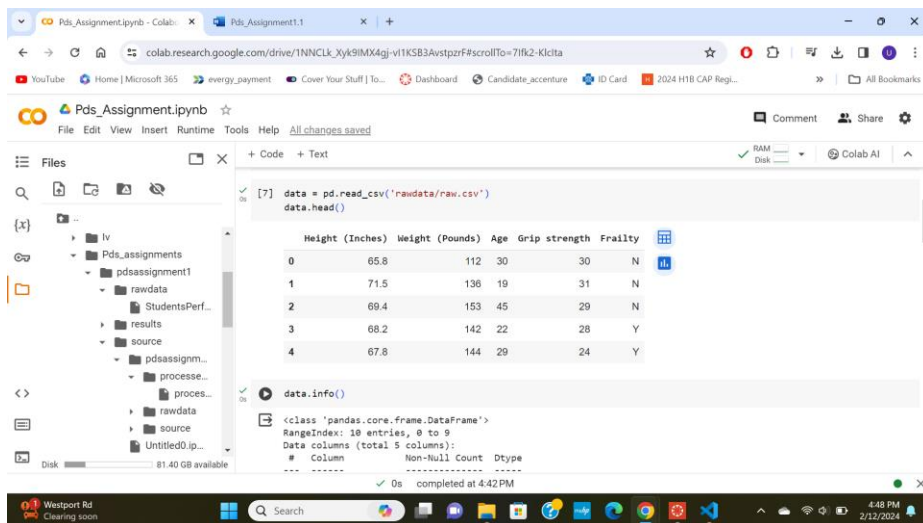
Mounted at /content/drive

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

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

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

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



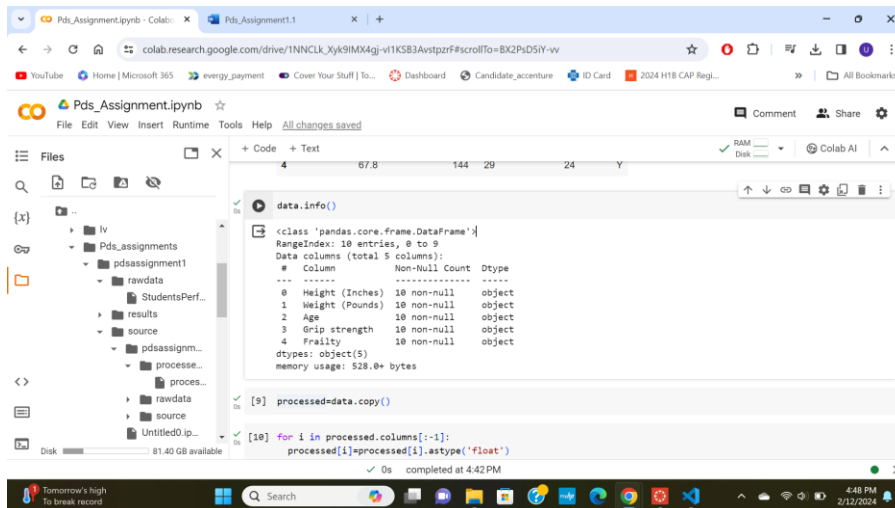
```
[7] data = pd.read_csv('rawdata/raw.csv')
data.head()
```

	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

```
[8] data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  -
0
```

Step 2: Processing of the data

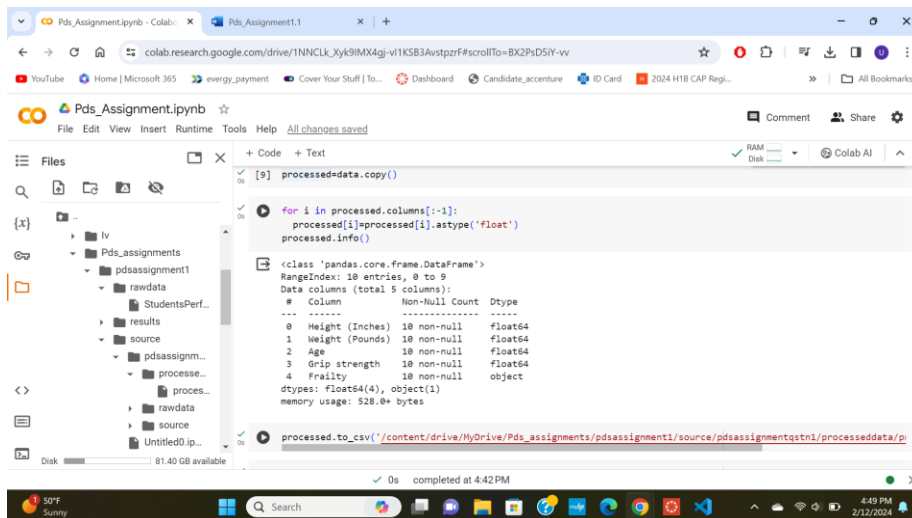


```
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 5 columns):
#   Column              Non-Null Count  Dtype
---  ---
0   Height (Inches)      10 non-null    object
1   Weight (Pounds)      10 non-null    object
2   Age                  10 non-null    object
3   Grip strength        10 non-null    object
4   Frailty              10 non-null    object
dtypes: object(5)
memory usage: 528.0+ bytes

[9] processed=data.copy()

[10] for i in processed.columns[:-1]:
      processed[i]=processed[i].astype('float')
```

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



```
[9] processed=data.copy()

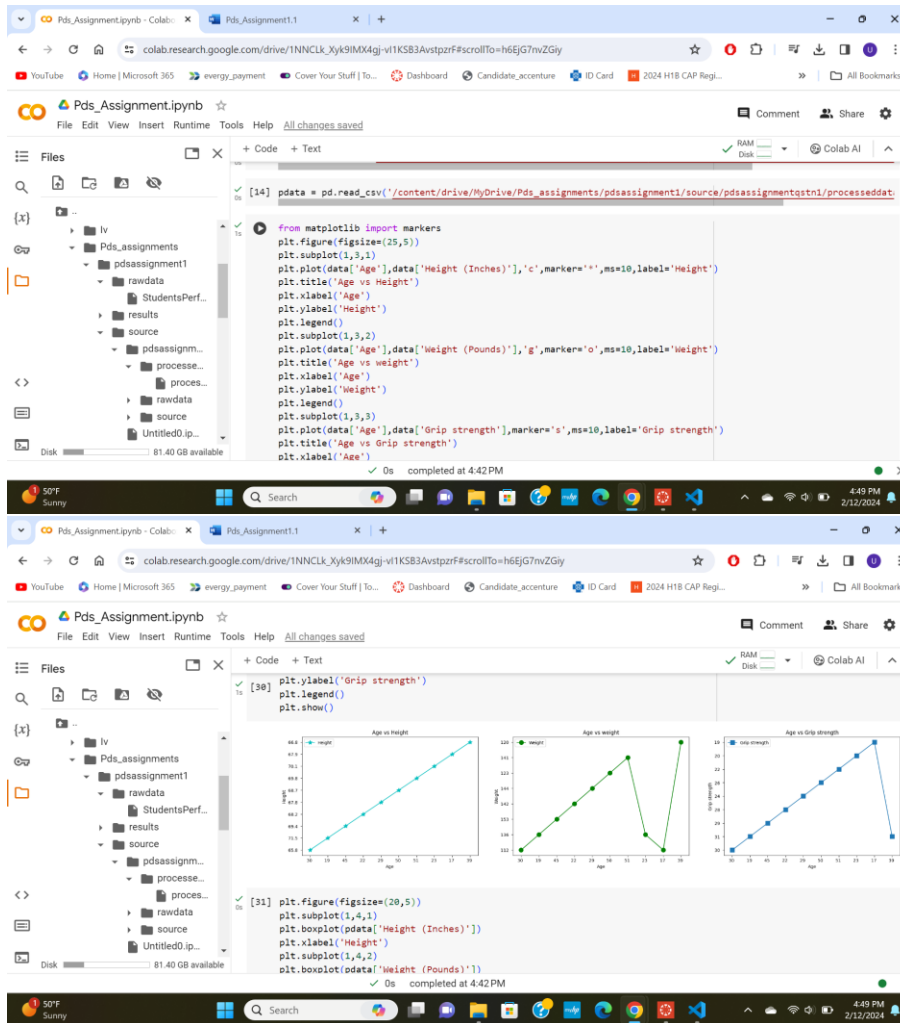
[10] for i in processed.columns[:-1]:
      processed[i]=processed[i].astype('float')
      processed.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 5 columns):
#   Column              Non-Null Count  Dtype
---  ---
0   Height (Inches)      10 non-null    float64
1   Weight (Pounds)      10 non-null    float64
2   Age                  10 non-null    float64
3   Grip strength        10 non-null    float64
4   Frailty              10 non-null    object
dtypes: float64(4), object(1)
memory usage: 528.0+ bytes

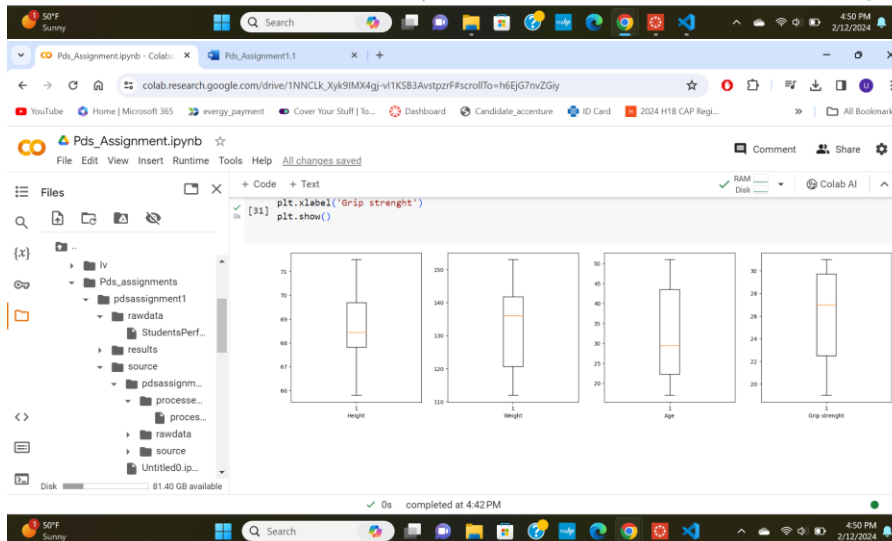
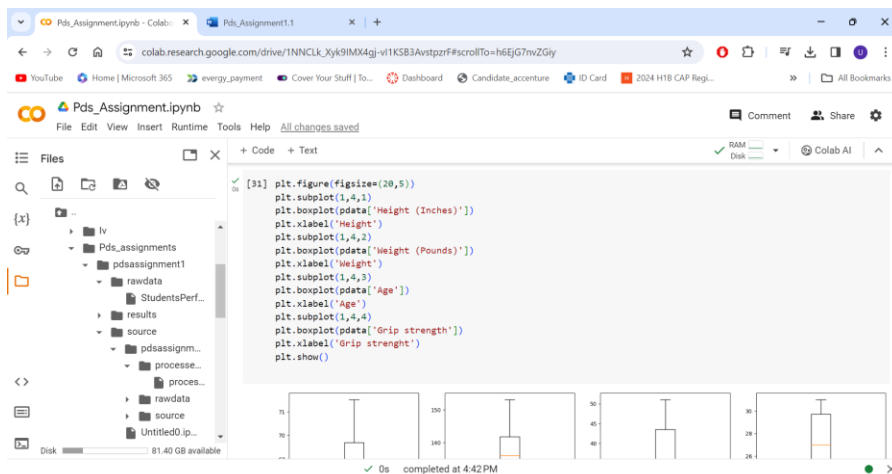
processed.to_csv('content/drive/MyDrive/Pds_assignments/pdsassignment1/source/pdsassignmentqstn1/processeddata/p
processed.csv')
```

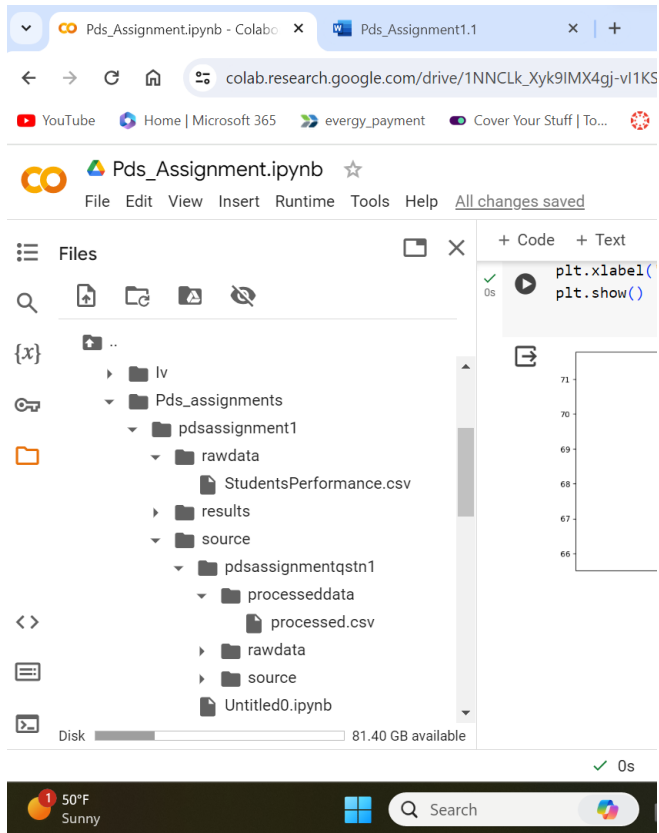
Post processing the data it is saved as processed dataset in processed.csv file under processeddata folder.

Step 3: Analysis of data



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





The above is the final file structure of processed data,rawa data,source,results.