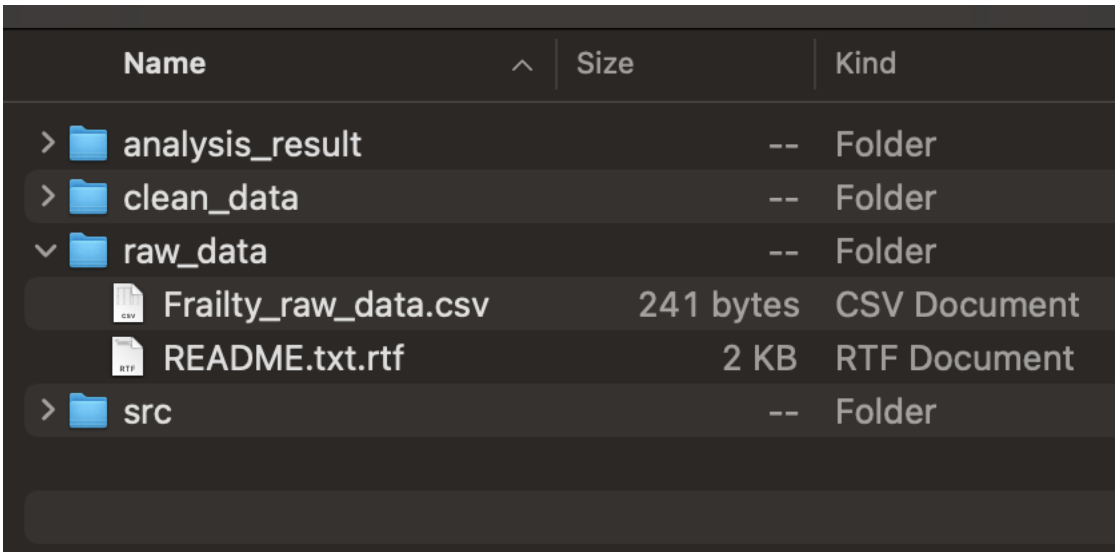
Assignment 1
Ambati Sai Aravind
16336124

1)

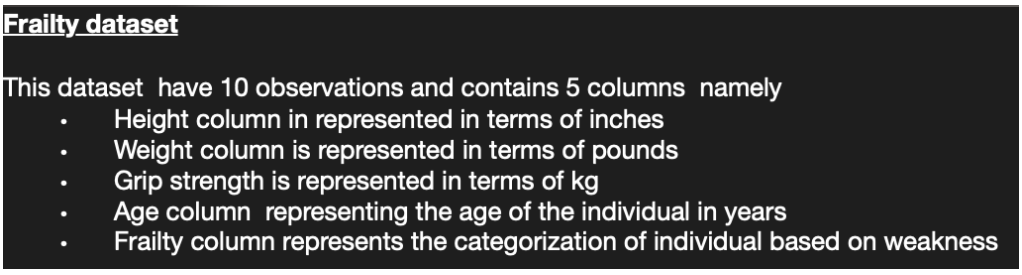
a) Data collection:

The data is collected and stored in a suitable format.

Folder structure:



- Gathered the data and stored it in a file with .csv format in order to perform the analysis.
- README.txt file contains the metadata about the dataset.

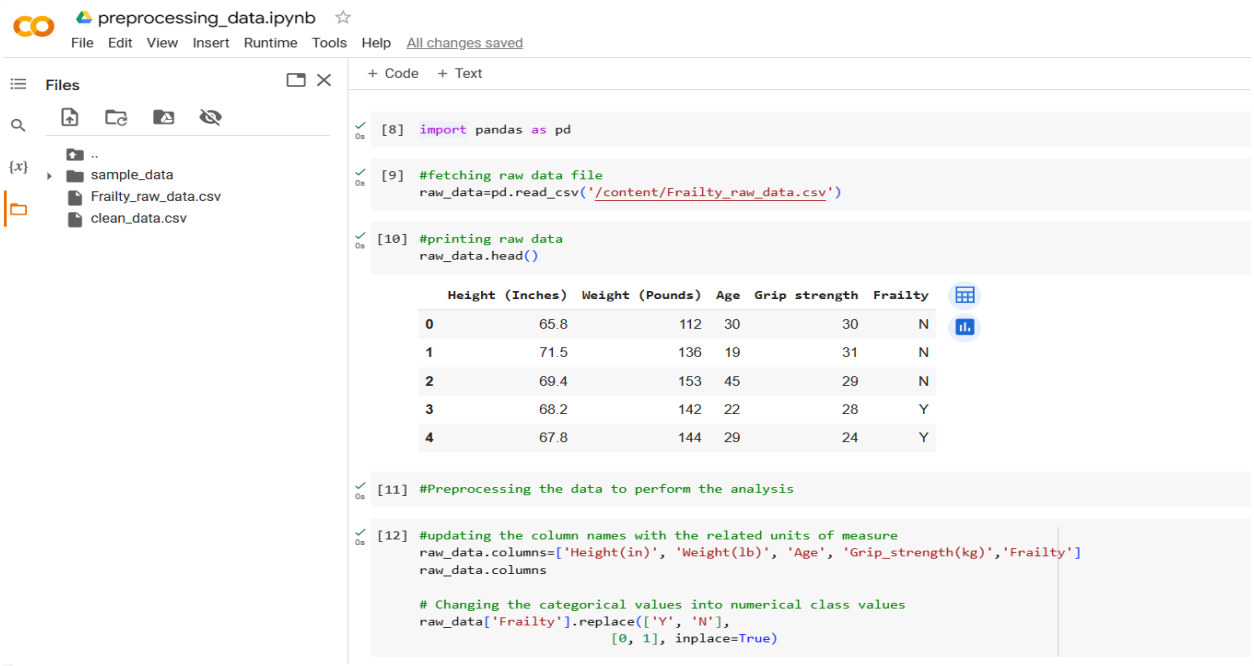


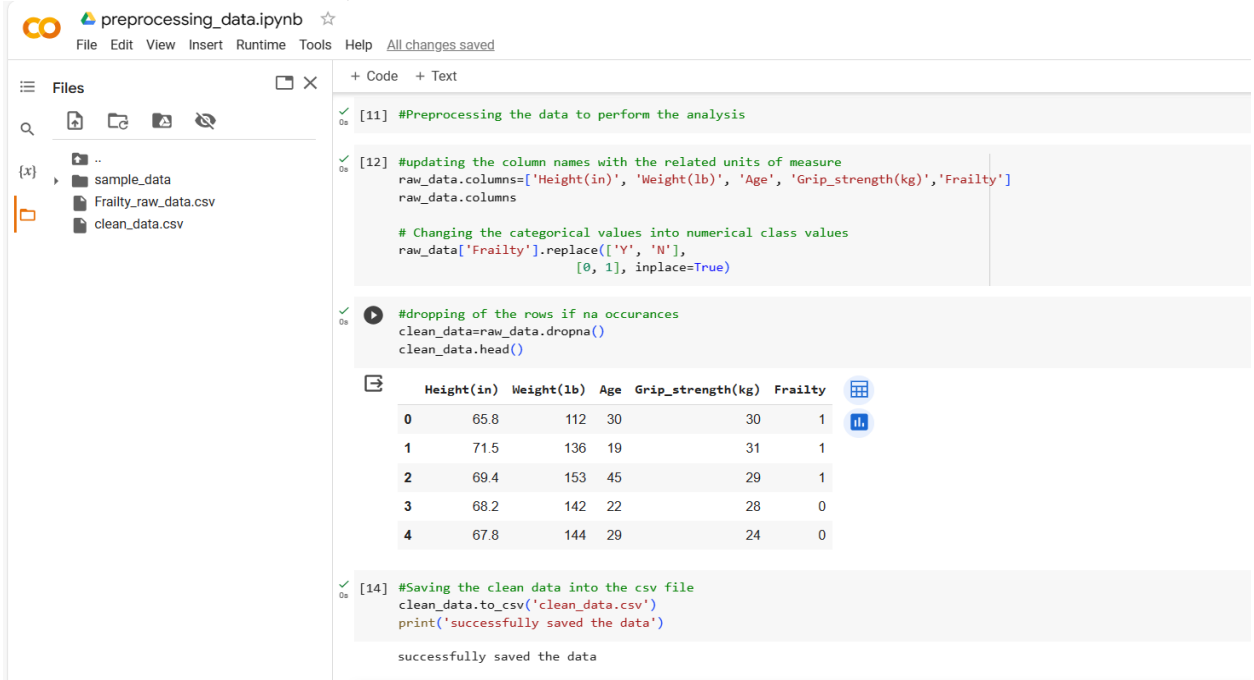
[README.txt file content]

Stage 2 – Data Processing:

Gathered data will be processed here in order to transform the data to be analysis ready.

- Preprocessing of data is performed using the scripts.





- Initially the data is loaded into the script and then certain preprocessing steps have been performed like renaming with suitable column names and converting the categorical values into numerical values making them more efficient for analysis. Validated if there are any records with NA occurrences. And the final obtained clean data which is ready to perform analysis is stored as a separate file.

Folder Structure:

Stage 3 – Data Analysis:

- Data analysis is performed by running the scripts to find the pattern of relation between the fields of the dataset.

analysis.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

+ Code + Text

```
[5] import pandas as pd
    import matplotlib.pyplot as plt
    import seaborn as sns
    from scipy import stats
```

```
[6] from plotly.tools import FigureFactory as FF
```

```
[7] data=pd.read_csv('/content/clean_data.csv')
```

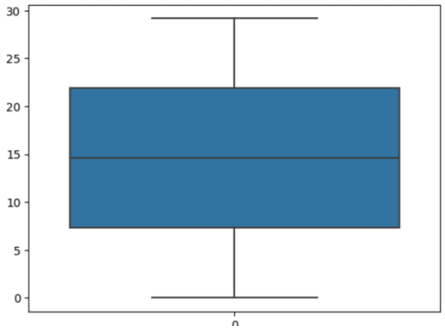
```
[8] data.head()
```

	Unnamed: 0	Height(in)	Weight(lb)	Age	Grip_strength(kg)	Frailty
0	0	65.8	112	30	30	1
1	1	71.5	136	19	31	1
2	2	69.4	153	45	29	1
3	3	68.2	142	22	28	0
4	4	67.8	144	29	24	0

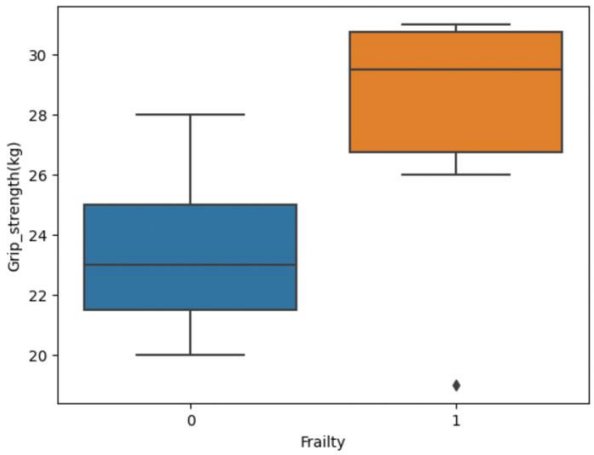
```
[9] #Performing the 2 sample t-test
    t2 = stats.ttest_ind(data['Weight(lb)'],data['Frailty'])
    print('analysis:',t2)
```

```
analysis: TtestResult(statistic=29.155389914737384, pvalue=1.3273210493059615e-16, df=18.0)
```

- **In this stage analysis is performed on the processed clean data.**
- **Two sample t-test analysis is done between weight and frailty and results are plotted.**



- **box plot graph is generated with the X-axis having Frailty and the Y-axis having Grip strength data.**



Analysis Result: The box plot representation clearly shows that greater the Grip Strength, the stronger will be the individual person.

Folder Structure: