# PRINCIPLES OF DATA SCIENCE ASSIGNMENT 1

**(Ujwala Dama 16334377)**

**STAGE1: DATA COLLECTION**

We have to collect the data in a suitable format.

• Acquired the data and put it in a.csv file so that the analysis could be done.

This dataset contains 5 columns

* height - in term of inches
* weight- pounds
* grip
* age
* frality - categorization of individuals

**STAGE2: DATA PROCESSING**

• Here, the collected data will be transformed so that it is ready for analysis. The scripts are used to perform preprocessing on the data.

The raw data will be cleaned and prepared in this phase.

Tasks to be completed

• Verify any missing data.

• Look for anomalies and errors.

• Add any metadata or annotations are required.

Folder organisation:

The Frailty Project

The data\_raw/

The file raw\_Frailty\_data.csv

Data cleanup

• raw\_Fraility\_data.csv

A screenshot of a chat

Description automatically generated

A screenshot of a computer

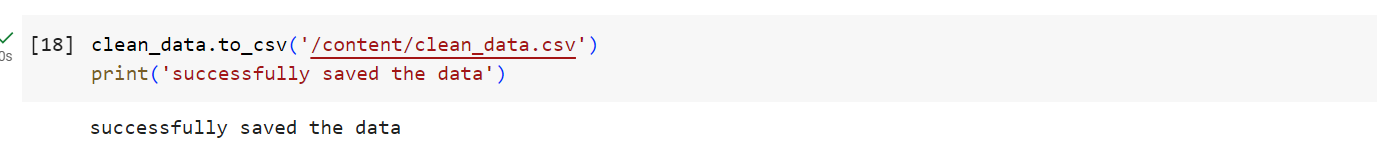
Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

In order to make the categorical values more useful for analysis, preprocessing measures have been taken, such as renaming with appropriate column names and turning the data into numerical values. Checked to see whether any records contain NA occurrences. And the final, clean data that was gathered and was prepared for analysis is kept in a separate file.

**STAGE 3: DATA ANALYSIS**

In order to glean insights from the data, we will do exploratory data analysis and produce visualisations in this step.

Tasks to be completed:

• Produce descriptive stats

• Produce visualisations to investigate the connections between various factors.

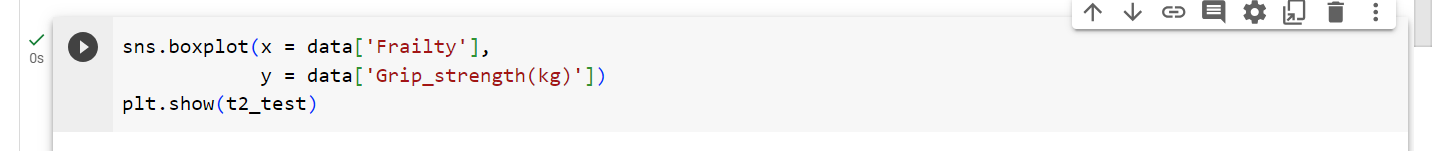
• Conduct any required statistical analysis.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated



• At this level, analysis is done on the cleaned, processed data.

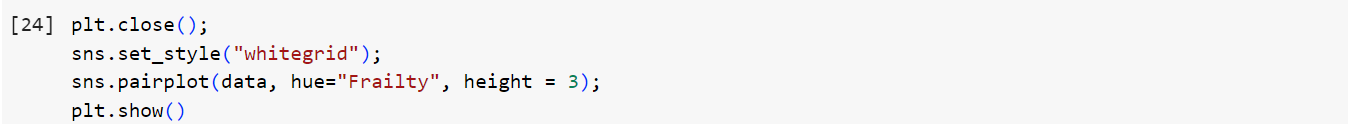
The findings of a two sample t-test investigation between weight and frailty are plotted.

• A box plot graph is created with data on grip strength on the Y-axis and frailty on the X-axis.

Analysis Finding: The box plot clearly illustrates that a person will be stronger if their grip strength is larger.

A screenshot of a graph

Description automatically generated



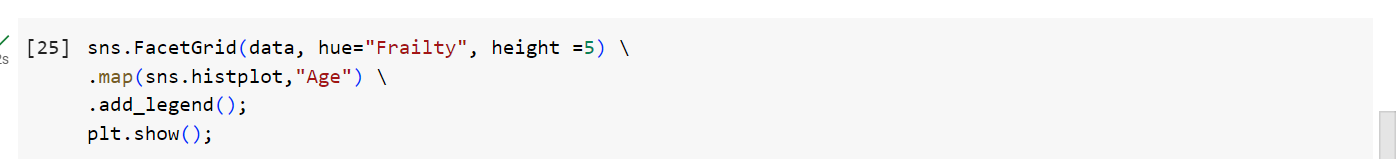
A screenshot of a graph

Description automatically generated

#Pairwise Scatter plot:Pair plot

#Observation:Here if we observe weight vs grip strength person with weight more than 135 and grip strength

.



A screenshot of a graph

Description automatically generated