Principles of Data Science Assignment 1
16325959

1. Data Acquisition:

-The first stage of most data-intensive workflows involves collecting raw data

-This data must be entered into a spreadsheet to be saved as a CSV file.

-The general convention is to put all project files in one directory, with one level of subdirectories for different file types such as data, source code, analysis results.

```
|-- strength_project
|   |-- raw_data
|   |   |-- frailty_data.csv
|   |   |-- README.txt
|   |-- data_clean
|   |-- results
|   |-- src
```

2. Data Processing:

-Once raw data is collected and placed in a project repository, it requires some processing or cleansing before it can be used in analysis.

-This step may include removing invalid data, refining the original data and removing outliers

-For example, after examining this particular dataset, we can improve the frail entry type for more consistent data representation

```
|-- strength_project
|   |—raw_data
|   |   |-- frailty_data.csv
|   |   |-- README.txt
|   |-- data_clean
|   |   |-- clean_frailty_data.csv
|   |-- results
|   |-- src
|   |   |-- clean_data.py
```

3. Data Analysis:

-Analyze and visualize the data for the frailty and grip strength in accordance to the height, weight and age.

```
|-- strength_project
|   |—raw_data
|   |   |-- frailty_data.csv
|   |   |-- README.txt
|   |-- data_clean
|   |   |-- clean_frailty_data.csv
|   |-- results
|   |   |-- test_results.txt
|   |-- src
|   |   |-- analysis.py
|   |   |-- clean_data.py
```