**Title:** Task B-1 Report

**Author:** Robin Findlay-Marks, s103603871

**Setup Attempt:**

Option B - Task 2: Data processing 1 took a long time to complete and had a lot of trouble understanding what I was meant to do.

The task seems to suggest that it is recommended for me to copy the code from P1 and then explain how it works. However P1 seems to do thing so differently from v0.1 that it is impossible without overhauling v0.1 to the point where it is basically P1 itself. I instead decided to figure out how to get the loading and processing done myself. I struggled for a while to get it to work properly but I’m fairly happy with what I did.

I made the program download a single dataframe, rather than two, which is defined with two dates. This dataframe is then split using one of three methods, with a third date, a ratio or simply randomly. I did this by having several smaller functions that first of all handle which splitting method to use, by checking which parameters are set, and second of all having functions dealing with the splitting by ratio or date.

The NaN issue was a annoying. I didn’t understand what I meant by the “NaN issue in the data” as I couldn’t find it referenced anywhere. I did see that something about NaNs was referenced in the P1 code, and after I realised that I researched what NaNs are and used the line from P1 to remove them.

Task 2 – 1.e was very confusing, and I’m not sure if it’s implemented or not. It seems to want me to scale the feature columns and store them as data structure, but it already scales the feature colunms and stores them as an array, which is a data structure. So as far as I can tell, this is already done, but that doesn’t seem right.

The rest of the explanations of how my code works can be found as comments in the code.

I have included screenshots for easy access.

Parameters:

A screenshot of a computer program

Description automatically generated

Downloading and storing of data as file:

A screen shot of a computer program

Description automatically generated

Splitting data by date

A screen shot of a computer program

Description automatically generated

Splitting data by ratio

A screen shot of a computer program

Description automatically generated

Processing of parameters

A screen shot of a computer program

Description automatically generated