**Background**

Markets often submit their data in a wide format which is not useful when we want to create a harmonised dataset with multiple markets.

We need to transform the data from wide to long to make it useful.

You can use any programmatic method you see fit to achieve the desired result however here is a guide on how you might approach the task.

**Data cleaning and Transformation:**

* Step 1. Load the value and volume sheets in from the “Data Test” file.
* Step 2. For each value and volume dataset:
  + clean it up so that we are just left with the Price Sector, Reportable Owner, Brand House, and Date columns in the form "w/e 04/10/2020"
  + pivot the data from wide to long on the Date columns, creating the “Date” Column of type Date.
* Step 3. Combine the value and volume data together.

**Bonus Points if you can do the following:**

* Rename Reportable Owner to “Holding\_Company”.
* Create a “PS\_HC\_BH\_Group” variable which is a concatenation of Price\_Sector,Holding\_Company,Brand\_House with a "\_" separator.
* Calculate the total Value (“Value\_T”) and total volume (“Volume\_T”) for each Date.
* Calculate each Brand\_House share over time (Value/ Value\_T).
* Any "ITUK" in Holding\_Company should be converted to "IMB".

**Desired Data Structure:**

The final column names should look like: "Date","Price\_Sector","Holding\_Company","Brand\_House","PS\_HC\_BH\_Group","Value","Volume","Value\_T","Volume\_T","Value\_Share","Volume\_Share"

**Question:**

What is the share of value and share of volume of the Brand House where:

Price\_Sector is "Premium" and Holding\_Company is "IMB" and Brand\_House == "JPS" and Date is "2024-04-21"