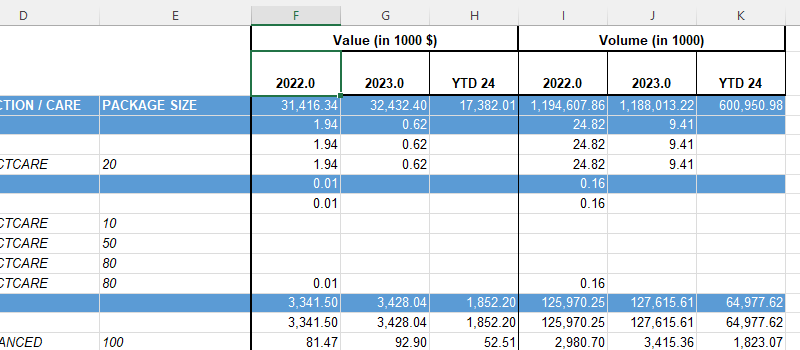
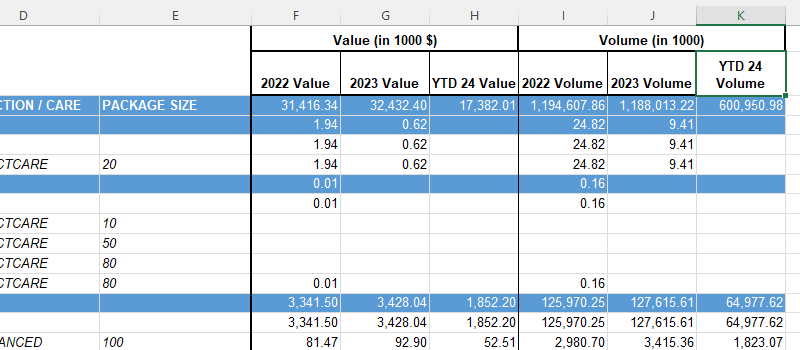
Data preprocessing

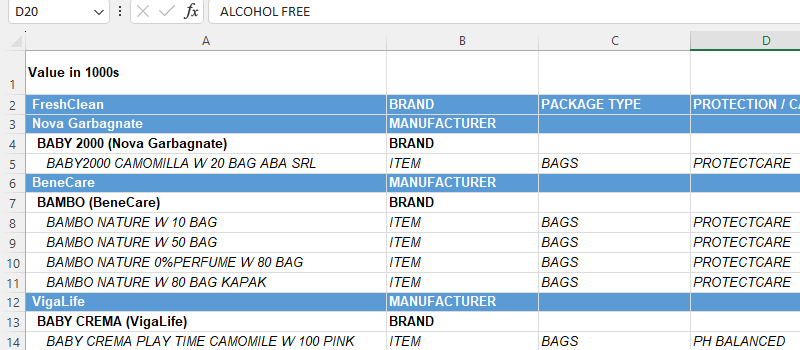
A first look at the data reveals that there is plenty to be desired in terms of the structure of the header of the source data. We need to fix that to obtain data that allows us to create a proper pivot table.



The presence of labels spanning two rows and merged cells representing multiple years pose a challenge to building a comprehensive pivot table report due to poor data structure. Specifically, when formulating the pivot table, we won't be able to differentiate between the 'value' and 'volume' for 2022, and the same issue would occur for other years. To ensure effective data analysis, it's essential that we restructure the data to enable clear and accurate distinctions.

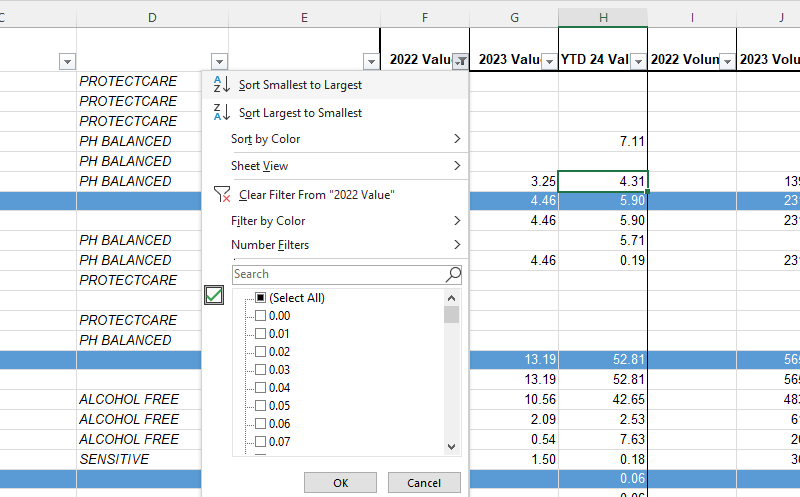


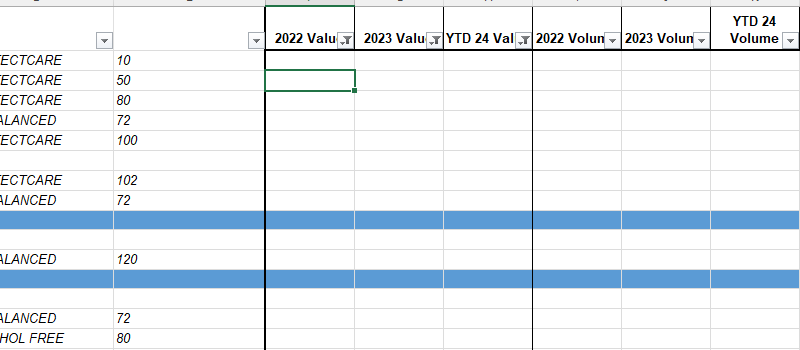
Therefore, the first step is to condense the information currently split across two rows into a single row. The oddly merged cell, which currently acts as a title, is also unnecessary.



If left in our source data, it could lead to complications when constructing the pivot table. Therefore, it's recommended to delete the first row, resulting in a singular header row that encompasses all needed information. This consolidated and organized header structure will better serve our data analysis needs.

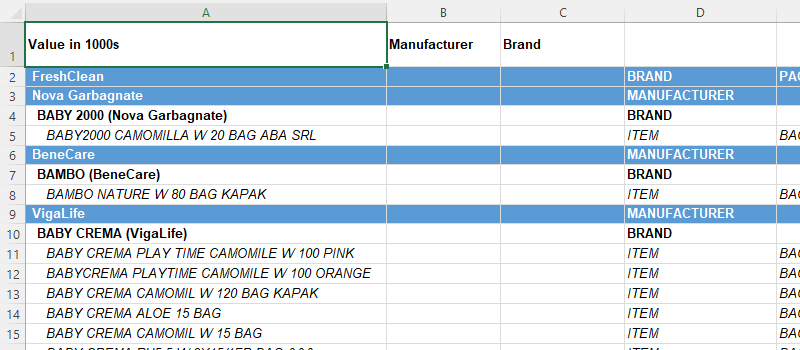
If there are products included in the source data that did not generate any sales during the period under analysis, they would result in empty rows in our report. These superfluous rows could clutter the pivot table and impede effective analysis. Therefore, it's prudent to filter out such null values, ensuring a clean and concise report.





We can delete these rows from the source table.

Another critical observation is that the source data doesn't present the manufacturer and brand information in a way that is suitable for pivot table analysis. When faced with this situation, it's essential to organize such information into distinct columns (where each row displays the specific product description, manufacturer, brand, etc.). Therefore, let's insert two blank columns titled *Manufacturer* and *Brand*. This adjustment will greatly facilitate our subsequent data analysis and report generation.

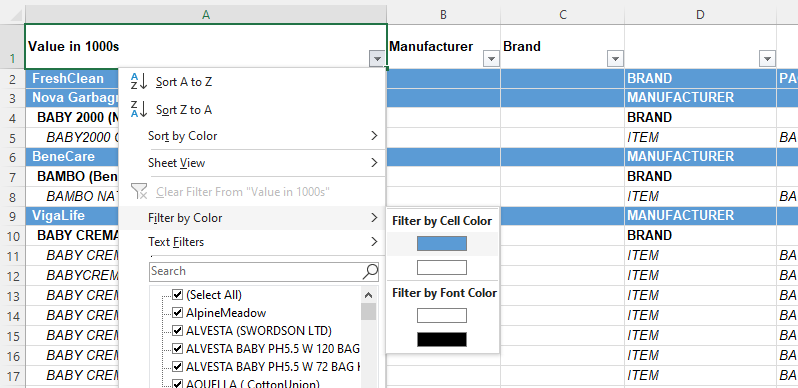


How do we fill in these columns?

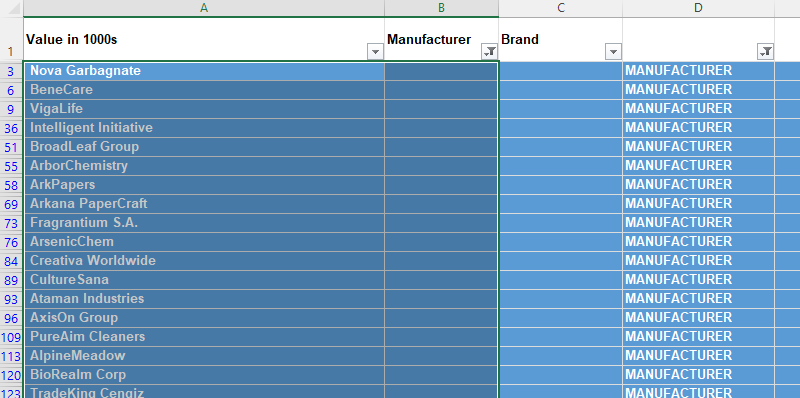
We can either do this manually, which would require more time and less sophistication.

The more elegant solution is to use Excel formulas, functions, and tools. If you decide to go with the second option, here is a description of the steps needed to do this efficiently.

During the initial exploration of the dataset, we found that all manufacturers' entries are colored blue. Let's apply a filter to isolate the blue rows. Alternatively, we could filter by the *Manufacturer* field in column D. However, it's worth demonstrating that color-based filtering can also be an effective method for data segmentation in certain scenarios.

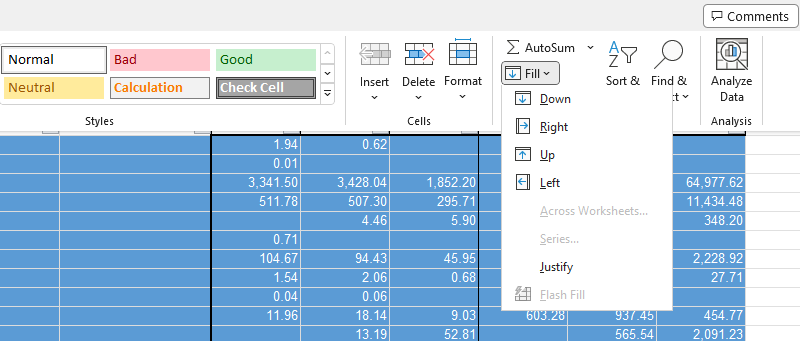


Following the filtering process, we'll have a list containing all manufacturers present in our data. Ensure to select the column that houses the manufacturers along with the adjacent empty column. This will prepare us for the next step of populating the new *Manufacturer* column.



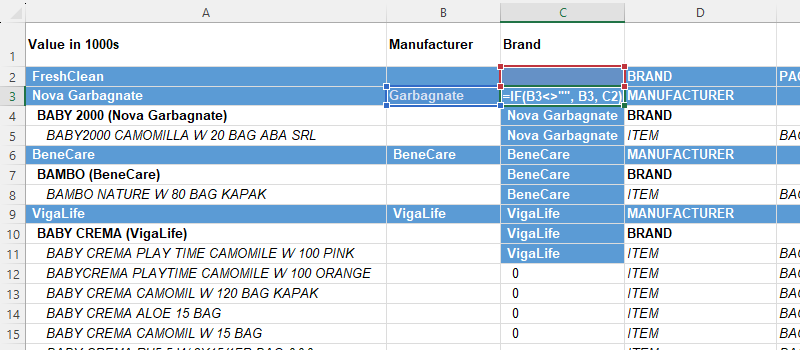
Copying values from column A to the *Manufacturer* column might not be straightforward because when you attempt to do so while your data is filtered, you may encounter a message stating that the action "won't work on multiple selections.

There is a nice workaround that is very useful in such situations. You can use the Fill functionality available in Excel’s Home tab. In this case, we need to use Fill Right.



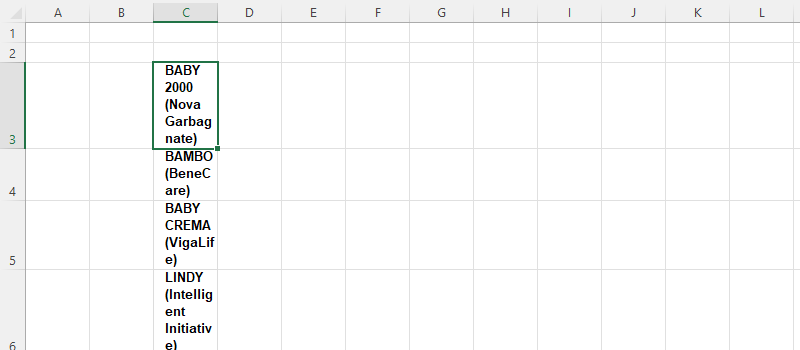
That’s good. Now you have the manufacturer name on every row in the filtered selection. However, please don’t forget that we need to have the manufacturer information in every row of the table, not just the filtered ones. To do that, we’ll need to use an IF function:

=IF(B3<>"", B3, C2)

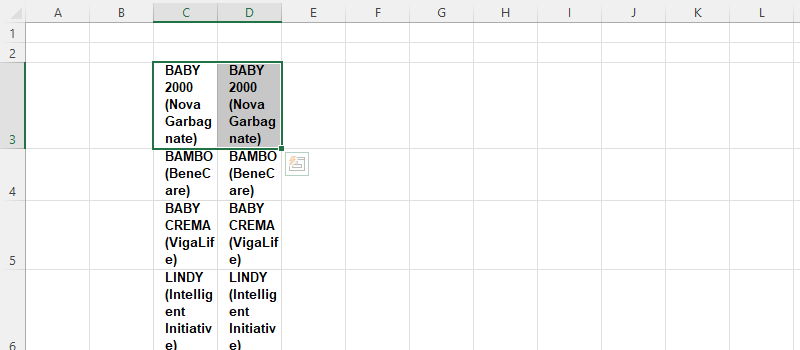


The function essentially checks whether the *Manufacturer* column contains a value. If it's not empty, it retains that value. If it is empty, it takes on the value from the row above. Developing such formula crafting skills is paramount for an analyst, as you'll often encounter similar situations when working with structured data. This formula allows you to extract the requisite manufacturer information on every row of the table. Be sure to copy all entries from column C to column B as values. You will need column C later to input the brand information.

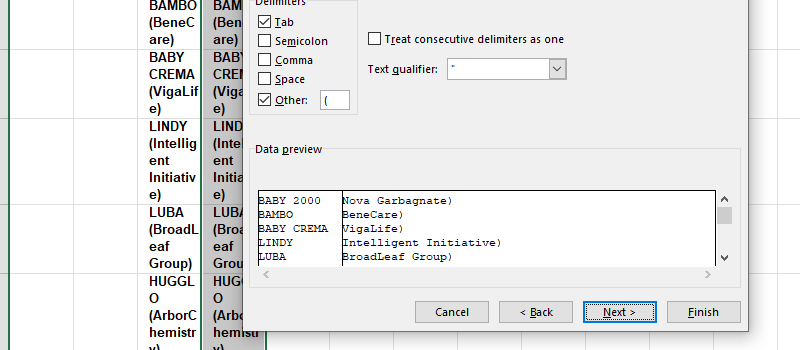
This is a bit more difficult to do. Filter for ‘brand’ in column D. Then you obtain a list with all brands in our source data. Copy and paste these names on a separate sheet.



Please note that besides that brand name we also have the manufacturer information in the same cell (in parentheses). We can use *Text-to-columns* to separate this information. However, we also need to preserve the current cell format to use it as a lookup value. Therefore, I recommend copying and pasting the column:

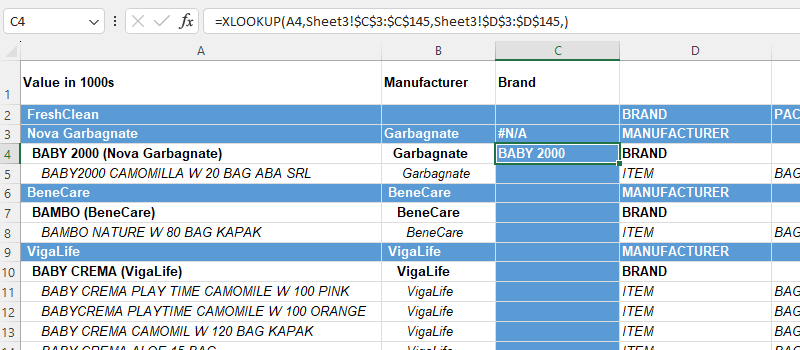


With the second column at our disposal, we can now utilize the *Text-to-columns* feature to segregate the *Manufacturer* information. Simply choose the *Other* delimiter option and input the parentheses symbol. This action will split the content of each cell into two separate cells at the point of the parentheses, allowing us to separate the Manufacturer information from the brand name.

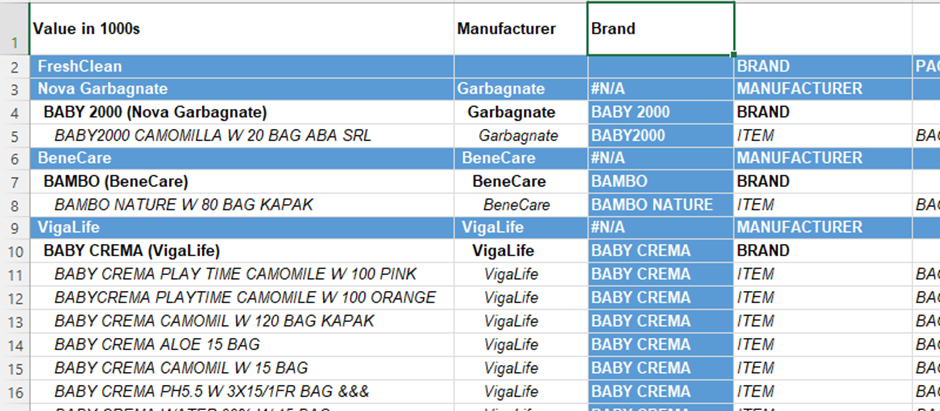


Next, we can retain the brand name and remove the manufacturer information.

Finally, we will employ the 'XLOOKUP' function to correspond the original information with the respective brand name.

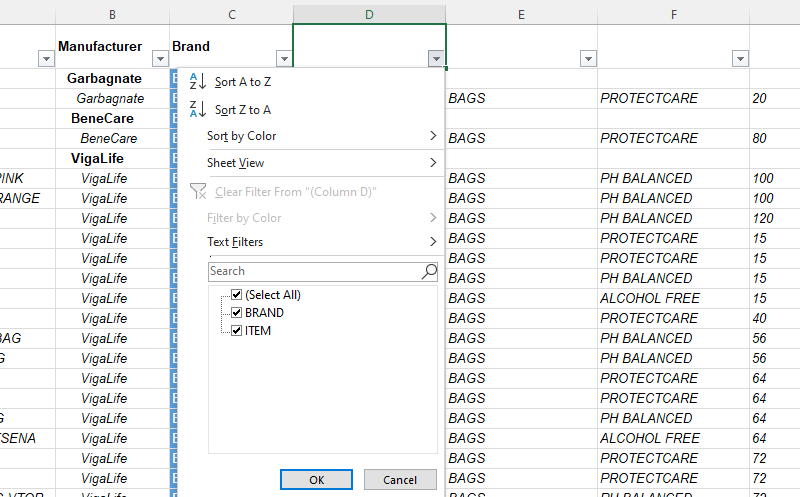


In this way, we can populate the *Brand*column. We can manually go through the source data and assign the respective brand. Or use formulas to do so.



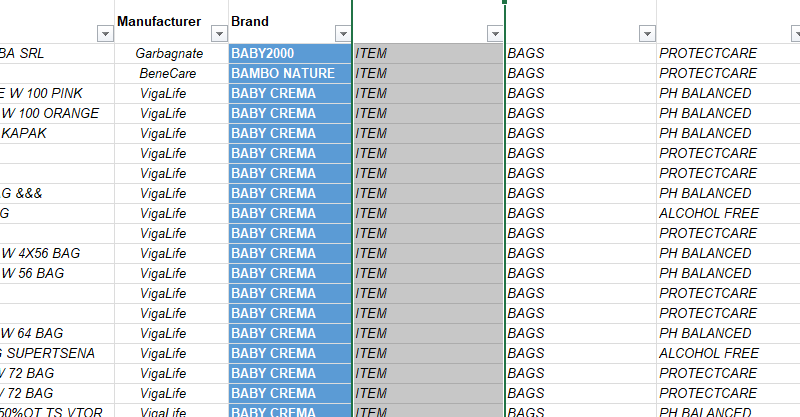
Once you have populated both the *Manufacturer* and *Brand* columns, you are ready to remove the subtotals at the brand level, which are represented by blue cells in our table. Why do we want to do this? Well, if we keep these subtotals, then our pivot table will double-count brand and manufacturer performance, wouldn’t it? This could lead to skewed results and inaccurate analysis, something we certainly want to avoid.

This time we can filter for ‘brand’:



Remove all subtotal blue rows. This is a very important step without which we cannot build a pivot table report.

Now that we removed *Brand*from the column in which we had a distinction between *Brand* and *Item*, we don’t have a reason to keep that column because it only says *Item*:



The preprocessing part of this project is challenging, but we are nearly there. The final step in this phase is to assign appropriate titles to each column in the header.

Let’s adopt the following column names:

Column A – *Product Description*

Column B - *Manufacturer*

Column C - *Brand*

Column D - *Package*

Column E – *Product Attributes*

Column F – *Size*

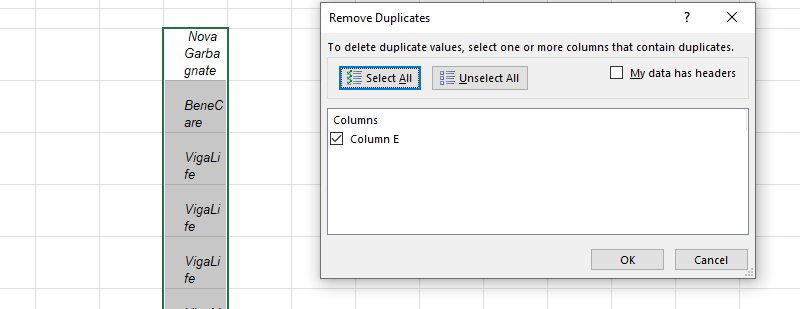
These titles offer a concise, clear understanding of what each column represents, which is vital for a smooth data analysis process.

Baby\_Care\_Task 1.xlsx

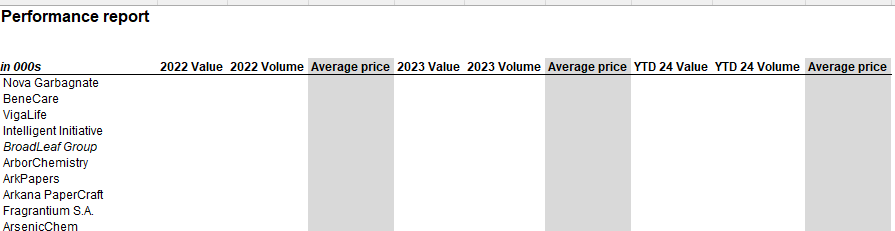
**Part 2:** Primary report table structure

In this part of the project, we need to create the structure of the primary report table. The goal is to display all manufacturers and their respective performances in 2022, 2023, and Year-to-Date 2024.

Start by copying the *Manufacturers*column in a new sheet and obtaining a unique list using the *Remove Duplicates* feature:



Once you have obtained the list, you can add the desired columns to the primary report table and apply professional formatting:

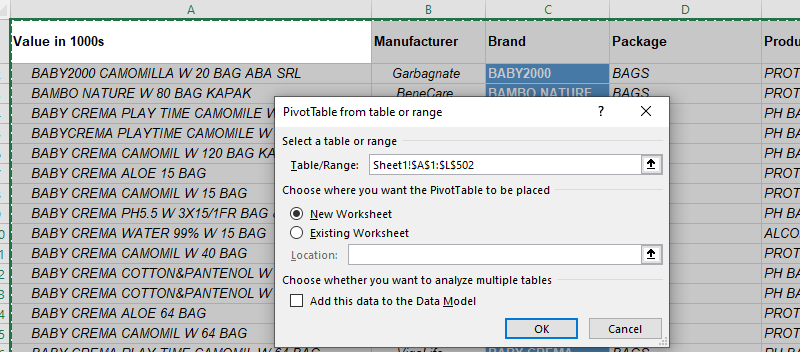


Baby\_Care\_Task 2.xlsx

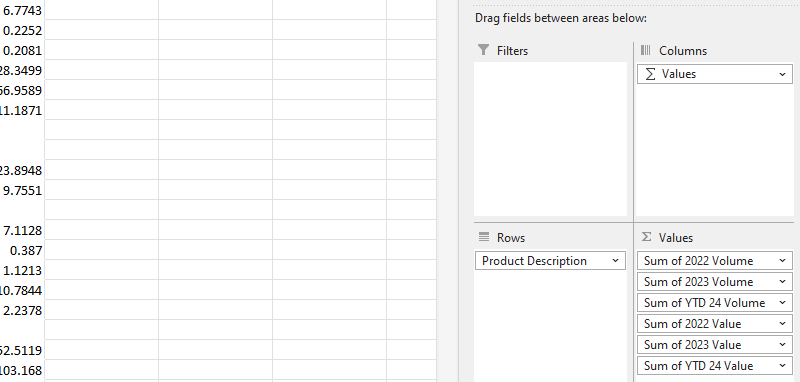
**Part 3:** Fill in the report

We didn’t have an easy time preprocessing the data, but all that work will help us create the pivot table that will serve as the source of the report we want to build.

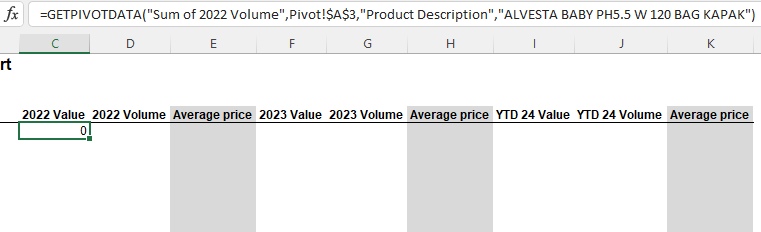
Select all input data and create a pivot table in a new worksheet:



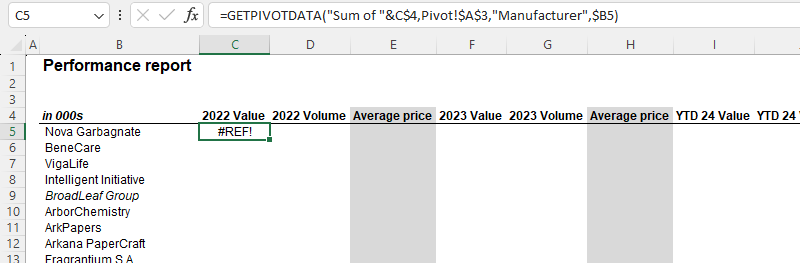
Here is how you can structure the fields in the pivot table:



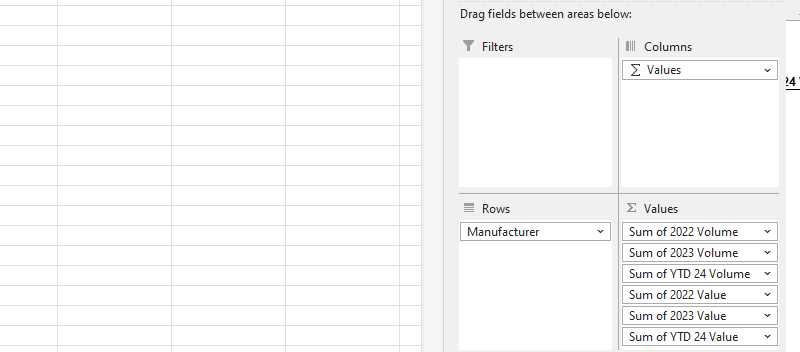
Next, you can link one of the cells inside your primary report table to the pivot table. In doing so, Excel will automatically generate a GETPIVOTDATA function, which provides detailed information about the specific value linked in the pivot table.



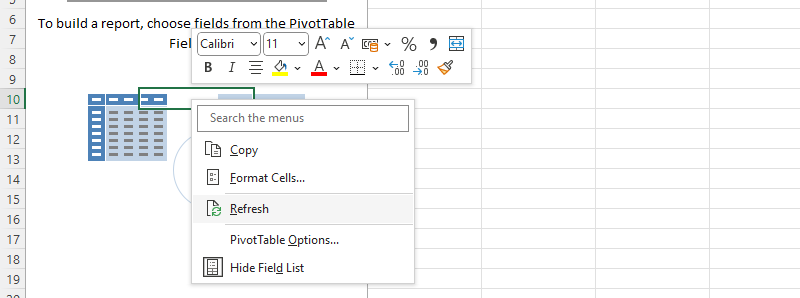
If we start modifying the formula, in theory it should work and update. However, there’s a crucial point to note. The pivot table will update as long as the data we ask to pull is available in the pivot table we have created. Not in the source data with which we have created the pivot, but in the pivot itself. This is very important. Please pay attention. If we adjust the pivot table and ask it to provide us a value that is not contained in it (but is contained in the source data), we will obtain a #REF error:



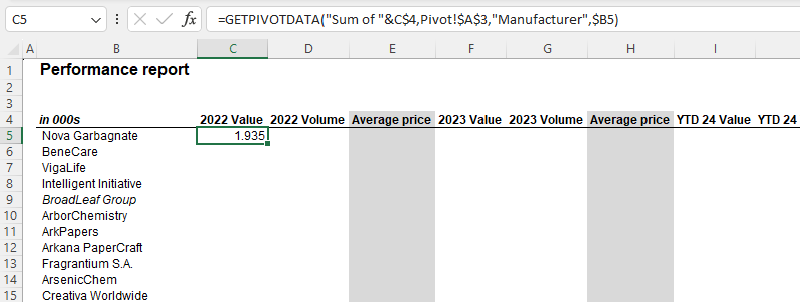
In this case, I’m asking the pivot table to return the sum of sales for the manufacturer *Nova Garbagnate*. The reason we get a #REF error is that the report contains product descriptions and not information at the Manufacturer level. From the start, I needed to create a pivot table considering this requirement. Let’s make the necessary change to update our pivot table:



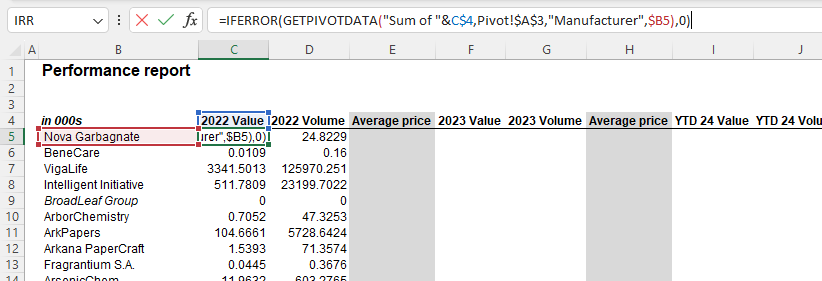
By substituting Product Description and Manufacturer we are able to Refresh the pivot table. Before pressing *Refresh* the structure and figures in the existing pivot table will not change.



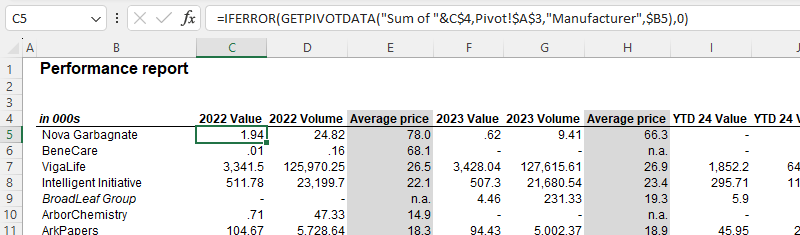
After we refresh the pivot table, the adjusted criteria inside the GETPIVOTDATA function will work. It is very important to fix cells properly, so that the GETPIVOTDATA can be pasted for all cells in the report.



Adding IFERROR in front of the GETPIVOTDATA function is another important touch. In some periods, there will be null values where we might have a #REF or #DIV/0 error. IFFERROR helps us prevent this issue, making the report easier to read.



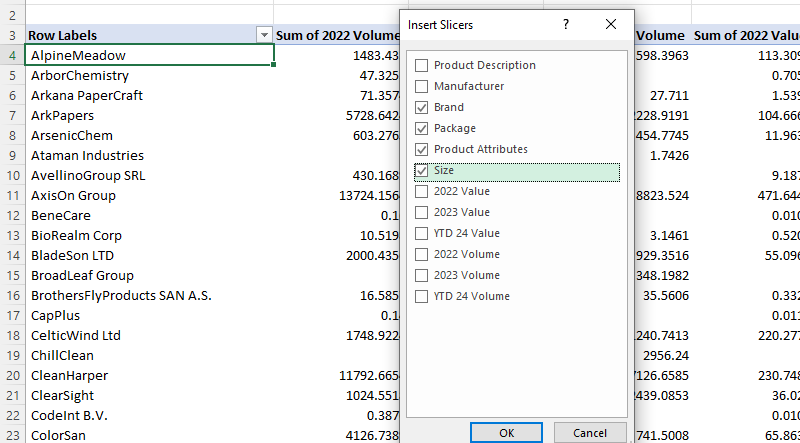
Make sure to format the numbers in the complete report, so that they look presentable:



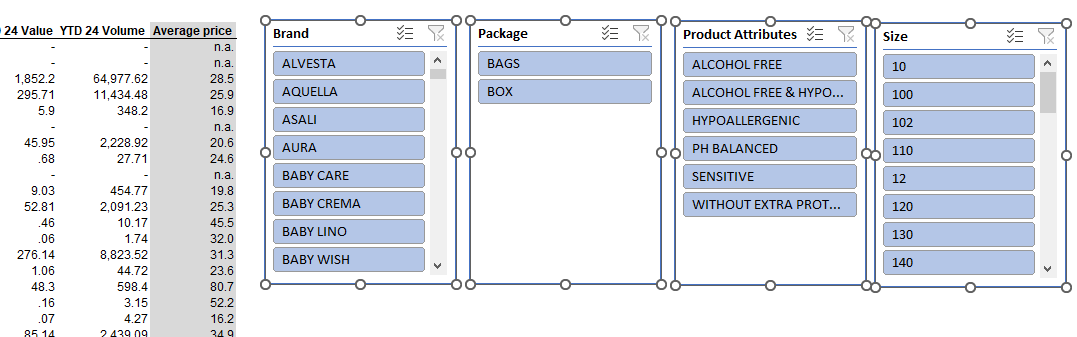
Baby\_Care\_Task 3.xlsx

**Part 4:** Insert slicers

We are almost there. The last piece of the puzzle is to insert slicers in our report. That’s fairly easy to do. Click on the existing pivot table and then opt to *Insert Slicers:*

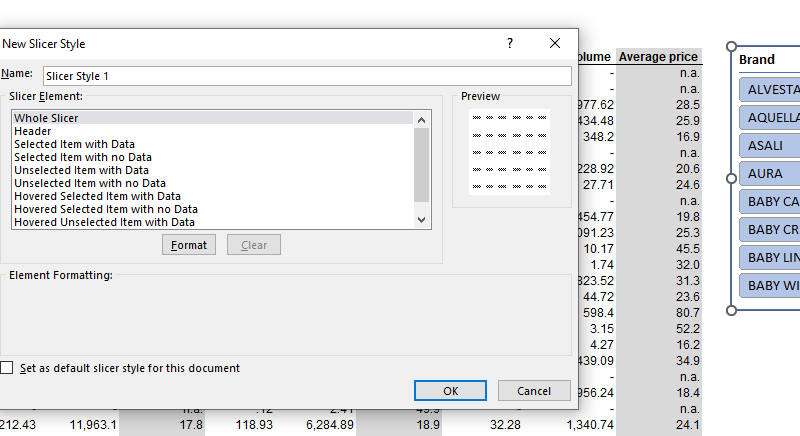
**

You can cut and copy the slicers to the primary report table sheet.

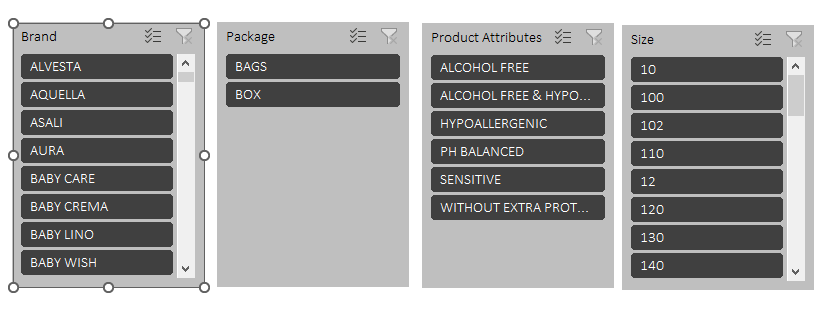


If we weren’t interested in formatting, we could stop here, but I personally prefer aligning the formatting of slicers to the formatting of the primary report table.

Select one of the slicers, and then the slicer menu will appear. Create a new slicer style:



Here is how my formatted slicers look like:



Baby\_Care\_Task 4.xlsx

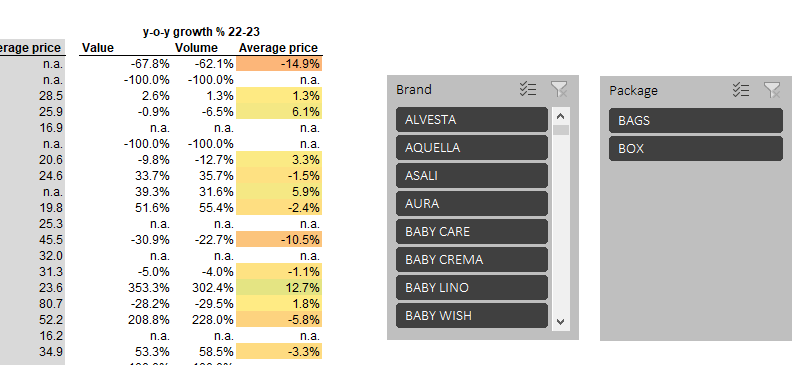
**Part 5:** Interpretation

Let’s calculate the growth rate of sales, volume, and average price for the period 2022-2023.

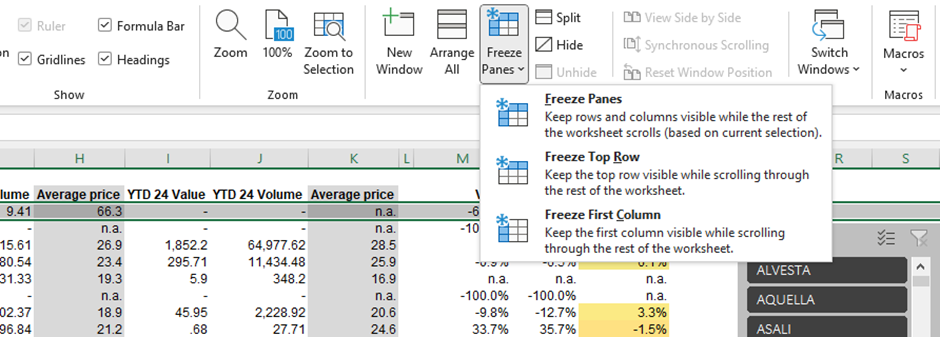


We divide the 2023 quantity by the 2022 quantity and subtract 1. As earlier, the IFERROR function helps us hide any potential errors that could be caused by missing values or no sales in the prior year.

Then we can use conditional formatting to highlight significant price deviations:



A final touch would be to fix the report header using freeze panes because we have many manufacturers, and the header gets lost if the user starts scrolling.



Let’s finally do some analysis!

Overall, the industry saw a 3.2% growth of sales year over year (22-23). This was primarily driven by a 2.2% increase of average prices and a 1% growth in terms of volume. Such moderate growth rates are not surprising as the baby care market is definitely in its mature stage. Moreover, the number of babies born in the region where the data was collected is very stable. In such environment, companies compete to gain market share at the expense of other brands.

**Top 3 companies by growth (> 1,000,000 units sold)**

1) ErusHealth Products – an astounding 761% boost of sales; average prices decreased -18% (which probably drove up demand), but slicers allow us to see that ErusHealth also introduced new products, which helped them grow revenue in this impressive way; And in fact, by filtering with the slicers we can see that they introduced alcohol products in 2023.

 2) More beauty – an amazing increase in revenue of 212%; In theory, when a company raises the price of its products, we should see a decline in volumes sold. However, this wasn’t the case here. More beauty managed to triplicate revenue by charging a bit more and expanding the quantity sold more than twice. The slicers show us that More beauty sold much less PH Balanced products and almost all growth came from Alcohol Free. If we use the Brand slicer, we will be able to see that More Beauty stopped relying on the Mors brand and instead all growth came from Wipest and more specifically from the introduction of the 120 size.

3) In third place, we have SigmaKappaZeta Co. with the impressive 175%. Their reduction in average price of -17% boosted volumes sold by 235%. In addition, the introduction of a new product size (120) and the very strong performance of the existing 72 PH balanced product contributed to the firm’s strong performance.

**Product attributes development**

When we slice *product attributes*, we can uncover some interesting insights:

1) Alcohol free products grew much faster than the rest of the pack (+20% in terms of sales and volume, average price stayed the same)

2) PH balanced products saw a -1.7% decrease of revenue despite a 5.2% price increase

3) Sensitive products grew +13% despite an 11% price decrease

4) Without extra protectcare indication products saw a -6% decline after a 2% increase in price

It is interesting to note that at the product attributes level price increases led to lower sales and the product categories, which kept their price stable and even reduced it experienced higher sales.

Therefore, one can conclude that a significant number of clients in this industry are very sensitive to price increases and there is an important price elasticity factor that needs to be considered by the top-level management of every firm.