# **Example 1**

Example 1, sourced from https://foresightbi.com.ng/microsoft-power-bi/dirty-data-samples-to-practice-on/, features a raw sales dataset that exhibits a typical structural inconsistency—multiple levels of headers. The raw dataset, provided in the repository as "Example 1 raw data," serves as a resource for practicing data preparation techniques.

The following steps can be applied to obtain a well-prepared dataset.

## **1.1: Changing the multiple headers structure**

To load the raw dataset into Power Query, transform it into a tabular format by using Ctrl+T, and then select "From Table/Range" from the Data menu.

A screenshot of a computer

Description automatically generated with medium confidence

**Figure 1.1: Raw dataset successfully loaded into Power Query.**

When multiple headers are identified for columns in a spreadsheet, the initial step involves identifying the columns that should remain unchanged. These columns can be referred to as fixed columns or anchor columns. In this particular example, the Order ID column is the sole anchor column that has been identified.

When working with multiple anchor columns, it is essential to merge them prior to advancing to the next step. The reasoning behind this will be further explained in the upcoming instructions. To merge two columns, select both columns (Ctrl + left mouse click), go to the Transform menu, opt for Merge columns, and then designate a separator for the merged categories from the ensuing pop-up menu.

The subsequent step involves performing table transposition. To transpose the table, navigate to the Transform menu and select Transpose. This operation reorganizes each row as a column and each column as a row. It is a crucial step in eliminating multiple headers associated with columns, as it transforms them into distinct rows. This highlights the importance of merging anchor columns prior to this operation. Failure to do so would result in remaining column headers originating from different anchor column rows. To provide clarity, I performed a slight modification to the dataset under analysis by adding an index column as an anchor column (Fig.1.2 and Fig.1.3).

A screenshot of a computer

Description automatically generated with medium confidenceA screenshot of a computer

Description automatically generated with medium confidence

**Figure 1.3: Modified dataset with index column after table transposition. Multiple headers are still present.**

**Figure 1.2: Modified raw dataset with index column**

As shown, multiple headers still persist in the table. However, in our specific example where only one anchor column exists, transposition can be directly applied, resulting in the Fig.1.4 outcome, where multiple headers are removed.

A screenshot of a computer

Description automatically generated with medium confidence

**Figure 1.4: Transposed table showing the removal of multiple headers.**

As the next step, the entries in Column 1 and Column 2 must be filled down, as the previous table organization was based on a different association of ship mode or segment with the values, whereas in our desired format, these entries are necessary. To accomplish this, right-click on the column header and choose the "Fill" option from the dropdown menu, then select "Down" in the pop-up menu.

As a matter of personal preference, I would opt to exclude the Grand Total column (last column). This is because the column itself represents a form of data analysis, and I prefer to maintain the prepared dataset without any pre-aggregated information, allowing for separate analysis steps. However, this decision is subjective and based on personal choice. To remove the Grand Total column, simply select the column with a left click, navigate to the Home menu, click on "Remove Columns," and choose "Remove Columns" from the subsequent pop-up menu.

The subsequent operation to carry out is column unpivoting, which involves transforming the column headers into row attributes. Before performing this operation, it is necessary to promote the first row to headers. This step is crucial because during the subsequent unpivoting process, we do not want to use "Column1," "Column2," and so on as row attributes for our observations. Instead, we want the "Order ID" to be recognized as a row attribute. To accomplish this, navigate to the Transform menu and select "Use First Row as Headers" in both the Transform menu and the pop-up menu.

The "Segment" and "Ship Mode" columns are already well-structured and do not require unpivoting. However, to unpivot the remaining columns, start by selecting the "Segment" and "Ship Mode" columns while holding down Ctrl and right-clicking. Then, navigate to the Transform menu, choose "Unpivot Columns", and select "Unpivot Other Columns" from the pop-up menu (result of this operation shown in Fig.1.5).

A screenshot of a computer

Description automatically generated with medium confidence

**Figure 1.5: Resulting table after the unpivoting operation.**

After renaming the column headers, the dataset can be loaded into Excel by selecting the Home menu, followed by the "Close and Load" option.

In the case of multiple anchor columns, the steps remain the same, with the addition of one extra step. After the unpivoting operation, it becomes necessary to split the merged anchor columns. To accomplish this, right-click on the merged column and select "Split Column" from the dropdown menu. Use the same separator that was employed during the previous merging operation as the splitting delimiter.