# Chapter 6 – Unpivoting Data Tables

## Topics:

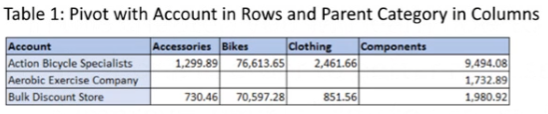
* Introduction to the Unpivot transformation
* Unpivot Columns
* Unpivot Other
* Handling Totals when unpivoting
* 2x2 Levels of Hierarchy
* Handling Subtotals

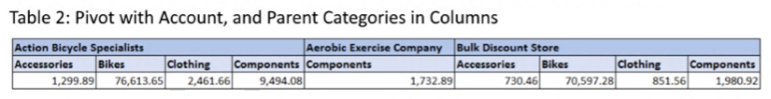
## Intro To Unpivot

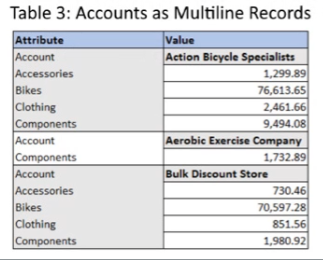
In this chapter, you will learn how to improve the table structure in your reports in order to better represent the data’s subject matter and improve [your] ability to analyze that data.

The Unpivot transformation is a cornerstone technique in addressing badly designed tables, and it is pivotal to your success as a data wrangler.

## Badly Designed Tables







All these tables are summary tables

* Rows represent more than one entity, or different entities
* Rows are not consistent in what they describe
* Columns are not consistent row to row

It is very difficult or impossible to report on this data in any other meaningful way (“**analysis locked**”)

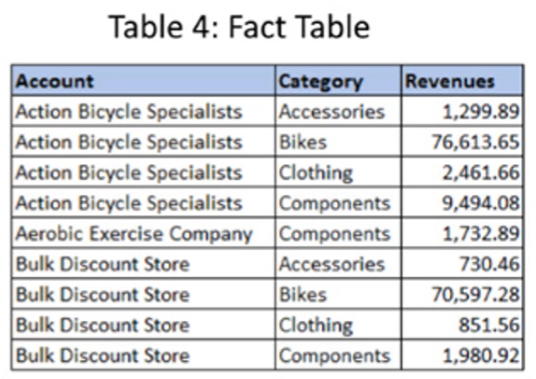
## Well Designed Table (Fact Table)

Fact tables allow for analysis (aggregation and summarizing

* Rows represent one consistent entity **Revenue for account by category**
* Columns are consistent and uniform

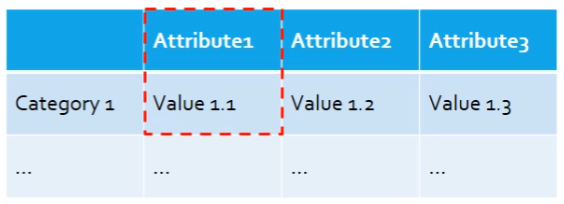
Tables in this format are the preferred data for BI reporting

Easily expand via merge and/or link to other data tables

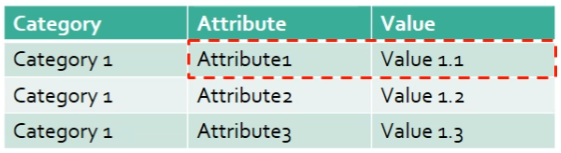


## Attribute/Value Pairs

Simple summary tables contain Categories in rows, columns which represent attributes, and values stored in the cells in each row.



Unpivoting these tables means creating a Category row for each of the associated Attribute/Value pairs



## Unpivoting Data Tables

In order to convert poorly formatted data, the Unpivot operation is used

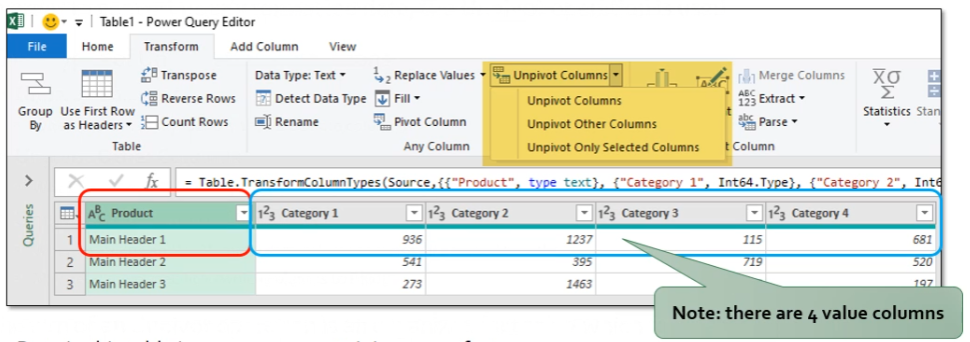
Three types of Unpivot transformations:

* Unpivot Columns
  + When you can easily identify the value data columns
* Unpivot Other Columns
  + When you can easily identify the nominal/categorical data columns
* Unpivot Only Selected Columns
  + When the Value and/or nominal data columns are not all grouped together
  + May need to be specific in which columns to Pivot

The aim of an Unpivot operation is an organized fact table which facilitates reporting

One entity per row, consistently described

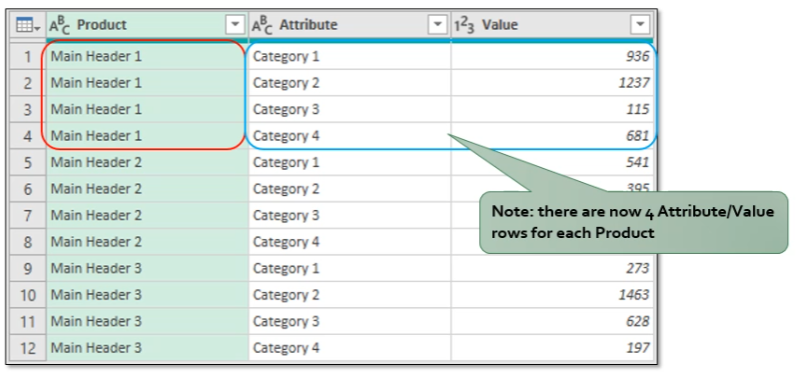
## Unpivot (other columns in this case)



Data in this table is a summary containing many facts per row:

* Product 1 sales in Category 1, Category 2, etc

## Unpivot Results



Unpivot operation creates a separate row for each Product/Category pair.

Each row in this table describes only one fact.

Data is preserved

## Unpivot Other vs Unpivot Column

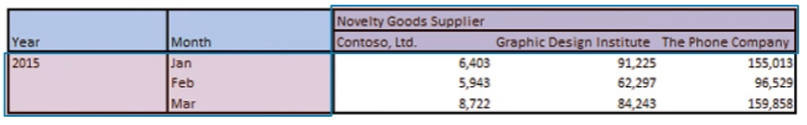
* The Categorical data (data that identifies the entity) are the axis you wish to unpivot on, the remaining data are converted to Attribute/Value pairs in the new table
* If you can readily identify the Categorical variable, you can **Unpivot Other Columns**.
* If it’s easily to identify the value data, you can Unpivot Column(s)
* The result will be the same
* The categorical data columns that you “unpivot” on are called Anchor columns
  + These stay in place (and are repeated) while the rest of the columns (un)pivot to the new positions

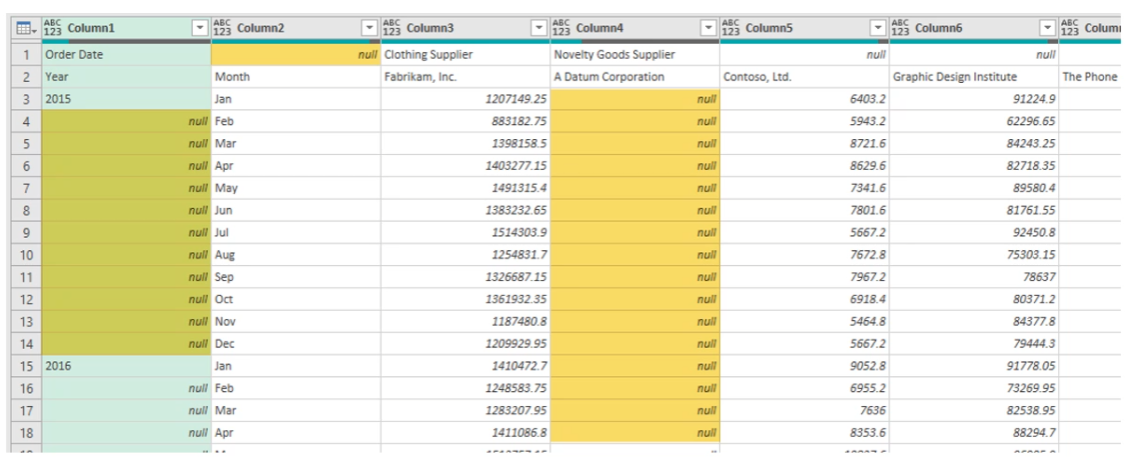
## Handling Totals

* Totals (and subtotals) in the data table are problematic
* Unpivoting a table containing total rows unpivots all source rows AND the total row
  + This **multiplies** the data in the resulting table (one extra attribute/value pair per category)
* Unpivoting a table containing totals column unpivots an extra attribute/value row
  + This **doubles** the value data in the table (all source fields + total itself)
* We don’t want either of these to occur
* Solution: **remove totals prior to unpivoting**
* Totals values can be easily obtained **afterwards** in reporting via Pivot Table or Power BI

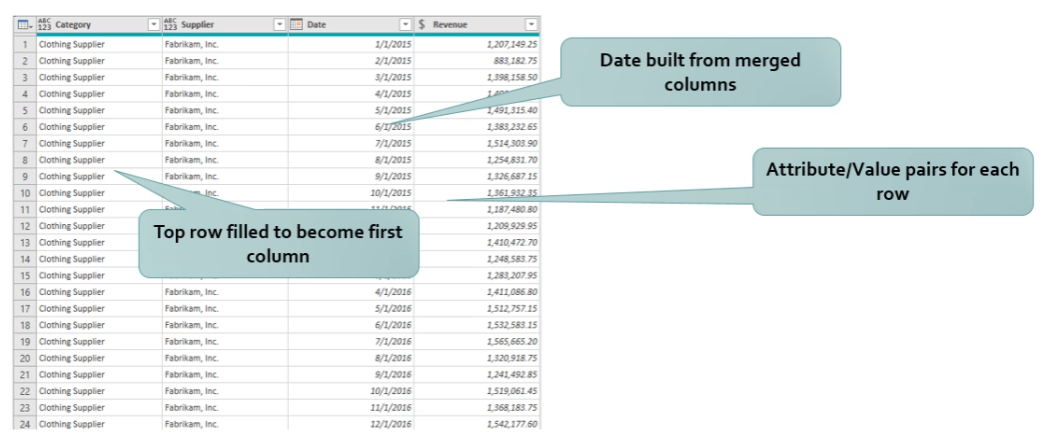
## 2X2 Levels of Hierarchy

* Excel data (very often) arrives as summary tables, rather than well formatted fact tables
  + Desktop Summary reports, etc
* Depending on the level of summarization, you may be able to simple unpivot to build your fact table, but often the data is multiple levels deep and obscured by labels (and decoration)
  + **Null** values in the Row and Columns cause issues
  + Overall category names as a column causes issues
* This hierarchy requires some extra effort to unpivot



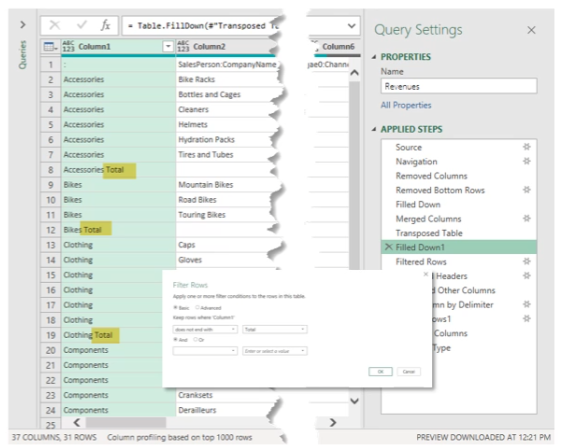


* **Fill Down** on the Year column
* **Merge** the Year and Month columns
* **Transpose** the table
* **Fill Down** on the Supplier Category
* **Apply First Row as Headers**
* Select the anchor columns and **Unpivot** the other columns



Handling Subtotals

* Subtotals In data tables add a little more complexity to the steps used previously
* Normally, you can identify these (especially from Excel) by the word **Total** in the cells
* Filters applied **at the proper stage** of the Unpivot operation can be used to easily remove these from the reporting data table
* Be careful not to filter too early: nulls get dropped by most text filter operations, so will be removed by a filter like “**does not end with**..”



## Summary

* Poorly formatted data tables cause issues with analysis via PivotTables and PowerBI
* The Unpivot transformation can be used to translate poorly formatted tables into easy-to-use fact tables
  + Unpivot Columns
  + Unpivot Other Columns
  + Unpivot Selected Columns
* Think of unpivoting as de-summarizing data tables
* Result is consistent data rows that facilitate analysis