# Preserving Context

## Intro

In this chapter you will:

* Learn how to preserve context from file names and worksheets when you append queries
* Learn how to apply the Drill down transformation on titles and dynamically preserve their content to tables
* Preserve titles as a pre-append phase by using Drill down and a custom column to combine multiple tables from a folder
* Preserve titles as a post-append phase by using Drill down, conditional columns, and the Fille down transformation to combine multiple worksheets from a workbook
* Learn advanced techniques for preserving specific cells from a grid using index and proximity cues

## Preserving Context

Append operations on source files have the potential to trim off useful data from the source

* Filename
* Category
* Source name etc

Taking steps to preserve this in the dataset, either pre-or post-append allows you to capture this data in the appended dataset, and use it for analysis

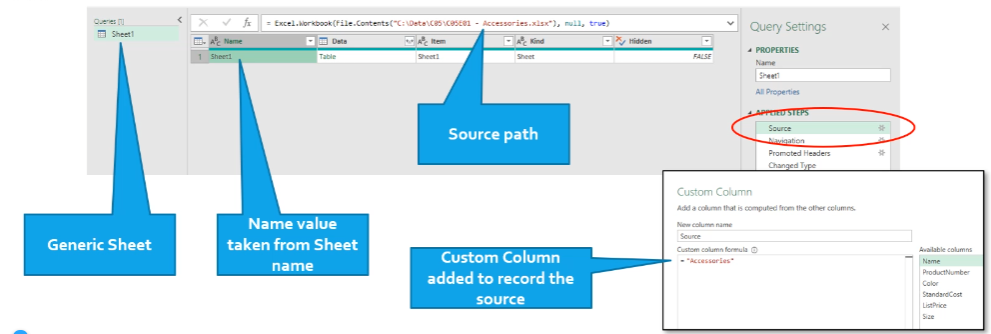
This data can be explicit (contained in the source content) or implicit(contained in the context of the source)

## Custom Column

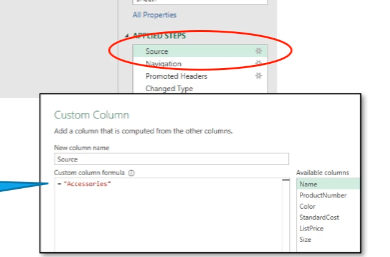
Often, the source name is obscured by generic object names, like worksheets, etc

The SOURCE statement in the query contains valuable data that could assist reporting

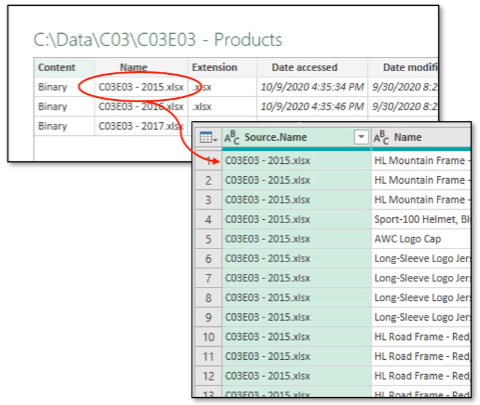
Custom Columns can be added to enrich the dataset (add columns or complexity) and capture the source data



The red part in the custom column says = accessories



## Filenames from Folder combine



When loading data from a Folder, the **Source.Name** field generated by the Combine operation can contain some useful context

This may be useful (or may not) depending on the name of the file(s)

**Custom columns** or **columns from Example** can capture this in the Sample Transform File steps

We’ve seen how to do this in Chapter 3

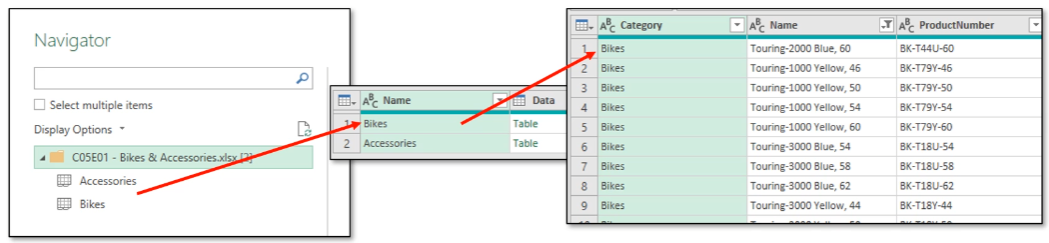
## Worksheet Names from Workbook

Worksheet names provide useful context for the datasource

The Workbook load operation provides the Name column in the preliminary load step

Exampanding the Table Data column while keeping the Name column preserves this context

This is only helpful if the worksheet names are meaningful.



## Before Appending

Certain preservation actions can be taken before the table Append operation

* Drill Down transformations to capture Worksheet Titles, for example

The text calls these pre-appended steps

These can be applied to single queries or to the Sample Transform File in a load from Folder operation

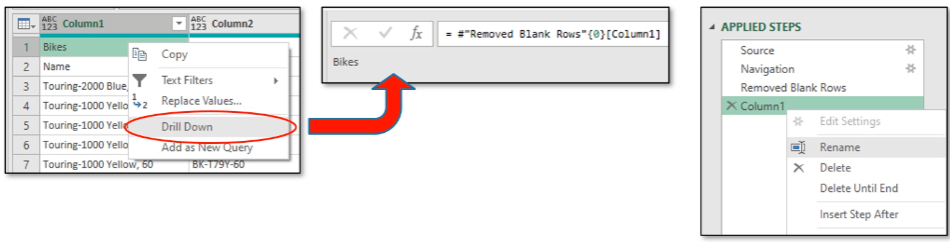
The result here is that each query table is adjusted to preserve available context before it is lost in the Append step Example anem of file ultimate source of that file

## Drill Down transformation

The Drill Down transform reduces the table row to a single cell value

Visually this is the negligible benefit

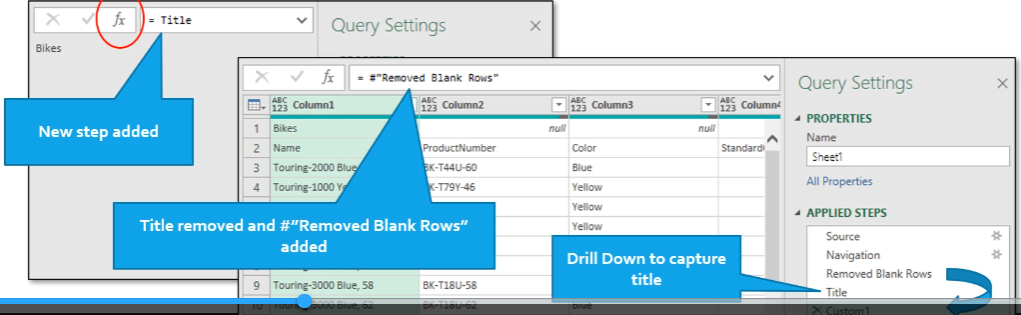
Procedurally, however, this provides a step that can be referenced later in the load script to record the results of the Drill Down step.



## Skipping steps in Applied Steps

The Add Function button can be used to add a step to the query procedure

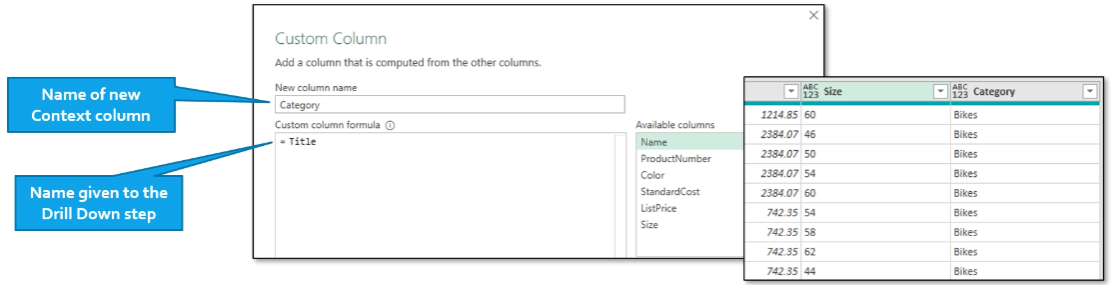
Referencing and earlier step by name will return the table state to that step, “undoing” changes made by Drill Down. This allows use to capture the results of the Drill Down, while not destroying the table.



## Reference Title by adding custom Column

After capturing the Worksheet title by using the Drill Down method, a Custom Column is then added to store the value of the Drill Down step.

* Add Custom Column
* Set its name to an appropriate value
* Set its formulate to reference the Drill down step (a single cell value!)



## After Appending

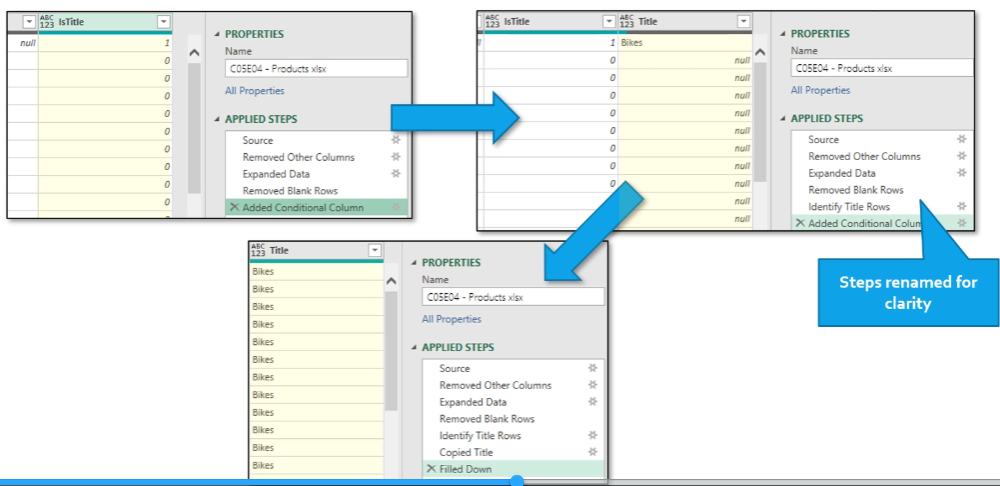
Note that the preceding operations may not be available to loading multiple Worksheets from a Workbook, as an example

Worksheet names may not have any contextual data to capture, or be meaningless

In these cases, changes can be made after the Append operation to capture context from worksheet titles

* After expanding, the data table is cleaned of blank rows
* Logic is used to determine the Title rows using a Calculated Column
* Another Calculated Column is used to transfer the title into the new column
* Fill down is used to fill the remainder of the column

Header promotion, etc. can proceed after this.



* Right click cell the click fill down and it will use that same value down the rest of the column

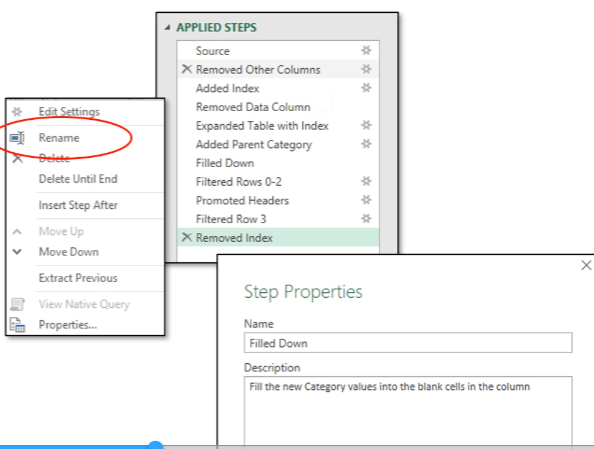
## Renaming Steps

It is often helpful to rename steps to indicate what the step doesn, or to provide a meaningful name to reference

* I.e. the Titlestep in the preceding examples

Rename a step by right-clicking on it and selecting Rename

Properties can also be added to provide a comment that displays when hovering on the step



## Using Context Cues

Context cues are structure elements in the data that can be used to help identify Titles and Context data

Many techniques (and more) are available

* Index Column (position on the sheet)
  + For example, title is always on row 2 of each file
* Anchor cells (position relative to a known cell value)
  + Search for known element and offset
  + M function **List.PositionOf**

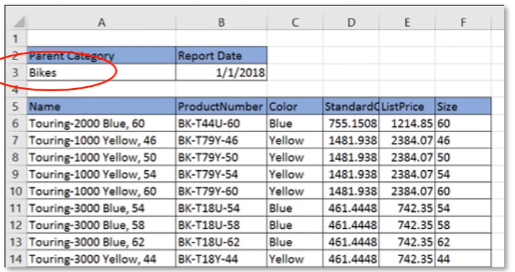
## Index Column

Known context position allows us to use an Index in the table data

Filtering out un-needed Index rows and operating on known Title row makes extracting the desired Context data straightforward

Conditional columns can respond to Index values to identify desired values

\* Index columns can also be used to find header rows in mismatched tables, facilitating the creation of Conversion Tables



## Cell Proximity

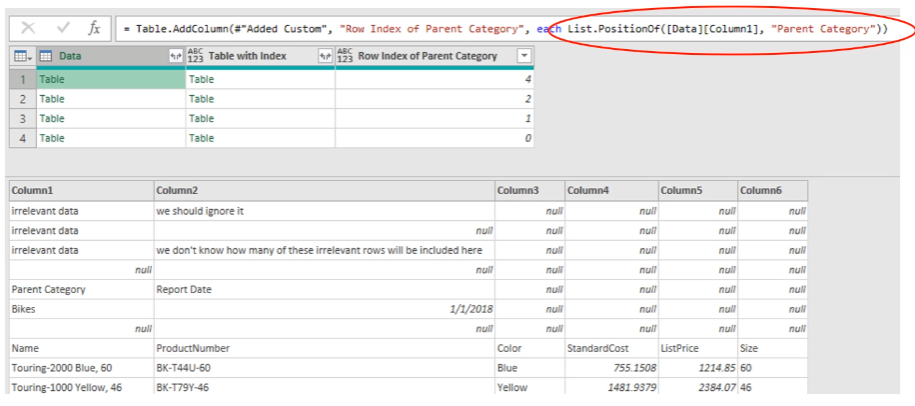
Often, worksheets are not uniformly formatted or are formatted differently when originating from different sources

In these cases, basic assumptions about position relative to the top of the sheet are not accurate

Instead, locating a common ‘anchor’ cell, usually a known heading, is possible using the **List.PositionOf** function in M

Additional detail can be placed in the preliminary load step to indicate the position of these anchor cells on each worksheet, and then scripting can continue from these values

## Cell Proximity - Finding Anchor Cells



## Summary

Append operations on data **can** capture useful context values from filenames, sheet names, and file paths associated with source files

This context information is **valuable** to help categorize the data and **enrich reporting capability**

This data can be lost during appending tables

Various methods can be used to capture this context:

* **Custom column** values entered manually
* Logic and **Conditional columns** to automate capture
* **Drill down** and **Fill down** operations to help automation
* Identification of position of context data to determine best strategy