# Chapter 5 Notes

## E5-1-1 Custom Column Technique

How do we extract metadata to gain insight?

We will be using the Bikes and accessories file

1. Import
   1. Lets import the accessories file by doing to data > get data > file > worksheet > accessories
   2. Then click on Sheet 1 and click load to > create connection
   3. Then go to data > get data < from file > from workbook >bikes
   4. Then click Sheet 1 and click load to > create connection
2. Rename
   1. Open the sheet 1 query
   2. To check where you imported the sheets from you can click on sheet1 in the Queries then look in the Query Settings > applied steps. Then click on Source and in the URL bar you will see the file path of where you got this data from.
   3. Rename the sheets to the proper names of the files
3. Combine
   1. Append Queries > Append as new then make the first one bike and the second one accessories
   2. Rename the new query to: Products
4. Custom Columns
   1. Now the Bikes and accessories are mixed but and we do not know which one is which.
   2. We need to go back to the Accessories and Bikes Query and add a custom column.
   3. Make the New column name: Parent Category
   4. Make the bikes have the value:

= “Bikes”

The accessories have the value:

= “Accessories”

1. However if we have many files like this this can cause a problem. So we will look next into how to do this automatically with many files.

## E5-1-2 Handling Context

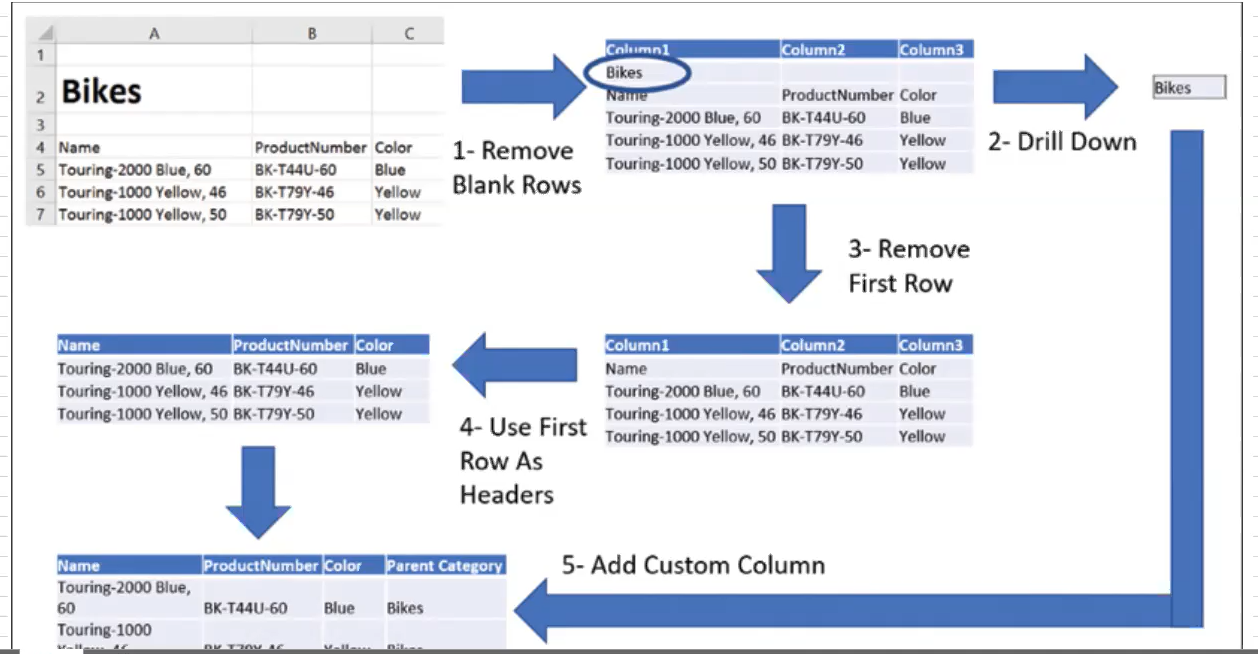
* To add context based on sheet names

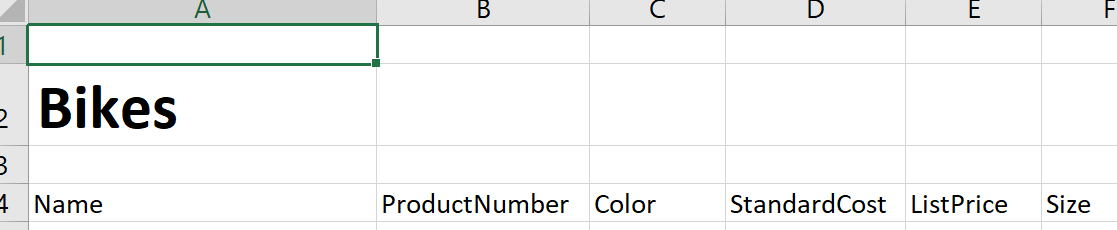
1. Data > Get Data > From File > From Workbook – we will choose the Bikes & Accessories file this time
2. Click on Accessories then Transform Data
3. Remove Changed Type and Promoted Headers and Navigation
4. Remove Item, Kind and Hidden
5. Now we will expand the Data object that has the data

* Make sure to uncheck the Use Original Column Names as prefix

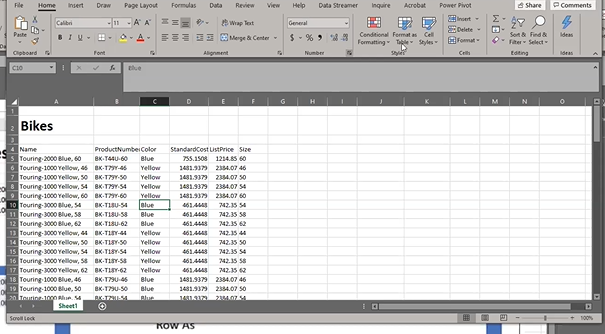
1. Promote First Row As Headers
2. Then in the Name Column just filter out the items that have the word Name
3. Change the First column from the title of Bikes to Category Name

## E5-2 – Keeping Titles Using Drill Down

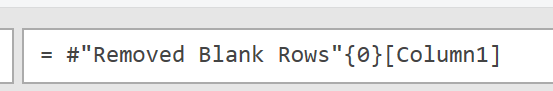




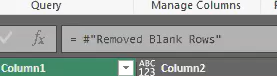
* Imagine your dealing with a file like this that has a weird title and spaces and not just the table



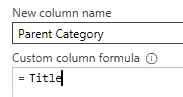
1. Go to Data> get data > from file > from workbook > bikes
2. Click Sheet 1 > load
3. Open up sheet 1 then remove from applied steps the: changed type and promoted headers.
4. Then go to Home tab > remove rows > remove blank rows
5. Click on the cell that contains the title value for example “Bikes”
6. Then right click and click drill down
7. Now it should look blank with just a formula in the formula bar that looks like:



1. Now in the Query settings – Applied steps rename the Column1 to Extract Title
2. Go back to your previous step above the Extract Title. It should be called Remove Blank Rows
3. Click on the step Removed Blank Rows then click, Click on the fx icon in the formula bar this new step should be after the Extract title so then we can skip that step and reference when the worksheet did display properly.
4. In the new step in the formula bar you can type in

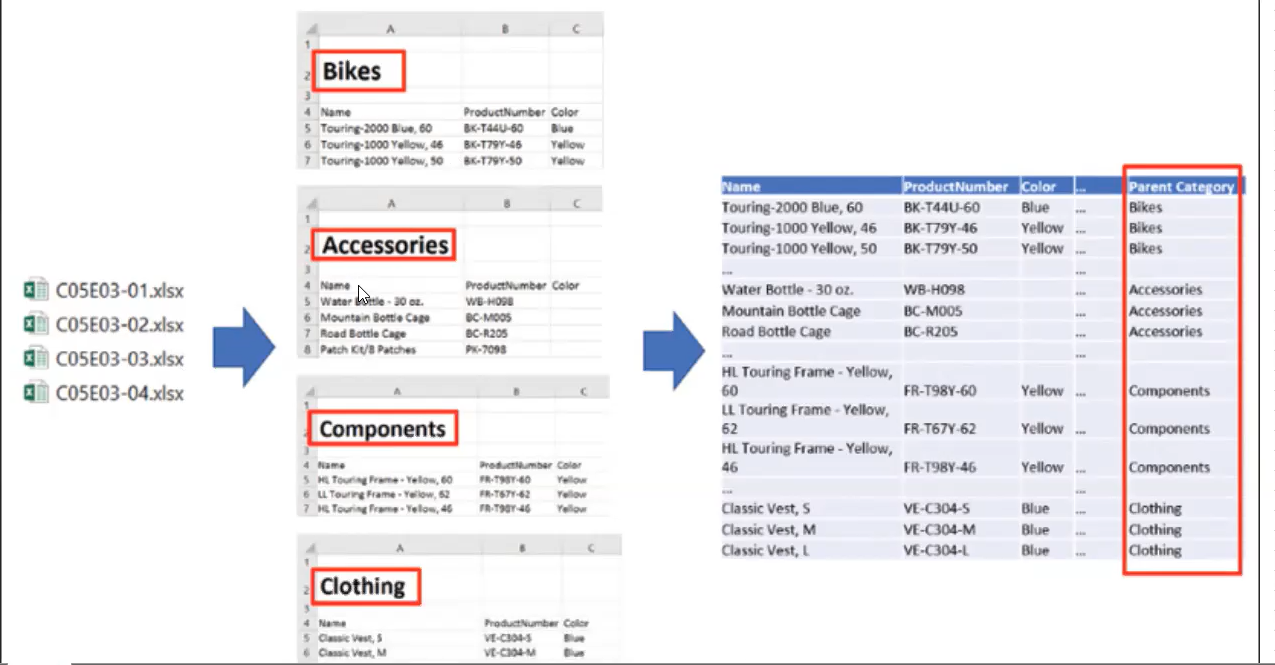


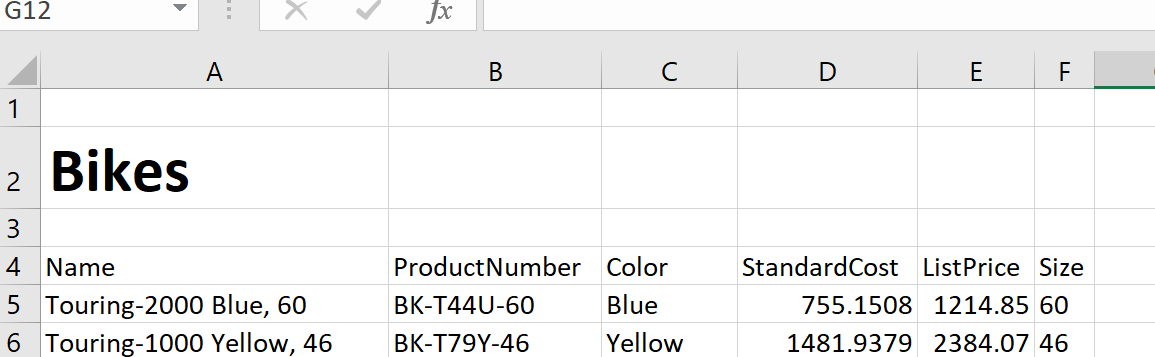
1. Now the worksheet should show the data
2. Now we will remove the first row. Go to Home tab > remove rows > remove top rows (and just type in the number 1 to remove only 1 row)
3. Then we will transform > promote first row as headers
4. Rename the Extract Title to Title
5. Rename the Custom1 (skipped step) to “Skip Step”
6. Click on the last step then go to Add Column > custom column
7. Call it “Parent Category” and in the formula we will type in:



1. Make first row as header

## E5-3 – Preserving Titles from a Folder



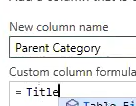


* Each file is written the same on the same lines
* Importing files from folders

1. Go to data > get data > from file > from folder
2. Click Combine and Transform data
3. Click sheet 1 then ok
4. Rename Transform Sample to Products Sample
5. In Products Sample remove the Promoted Headers
6. Then go to Home > remove rows > remove blank rows
7. Click on the cell that has the one title called “Bikes” then right click and select Drill Down
8. Now we want to add a extra step to skip it. Add a new custom by clicking the fx in the formula bar. This step should be after the Title step
9. Rename Column 1 as Title
10. Now in the new fx make reference to the Removed Blank Rows

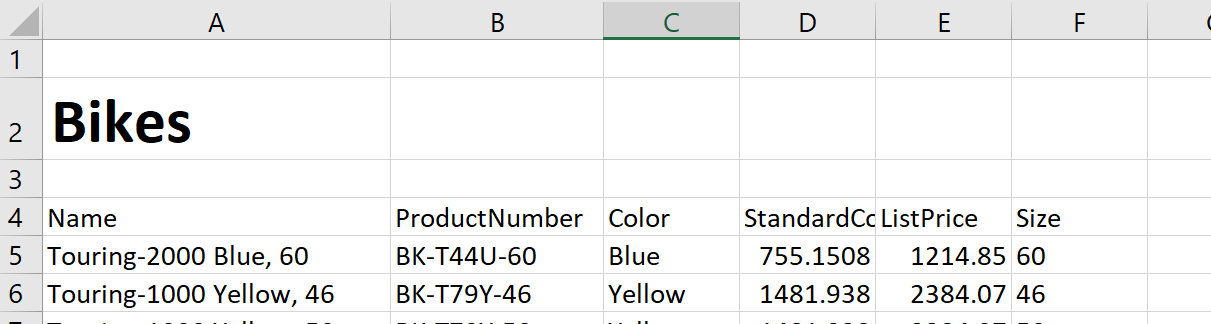


1. Now in Home > remove rows > remove top rows – when you do this just enter the number 1
2. Now promote the first row as headers
3. Add column > custom column
4. Call it “Parent Category” then make it equal to the Title

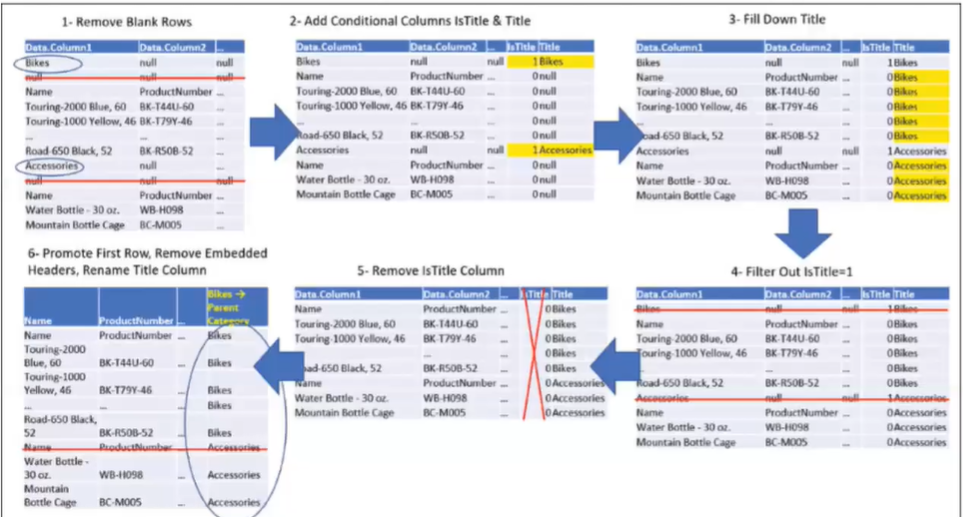


1. Click on C05E03 – Products and remove the Changed Type
2. Click on Products Sample and remove the Changed Type from the applied Steps to get rid of the errors row

## E5-4 – Preserving Titles From Sheets in Same File

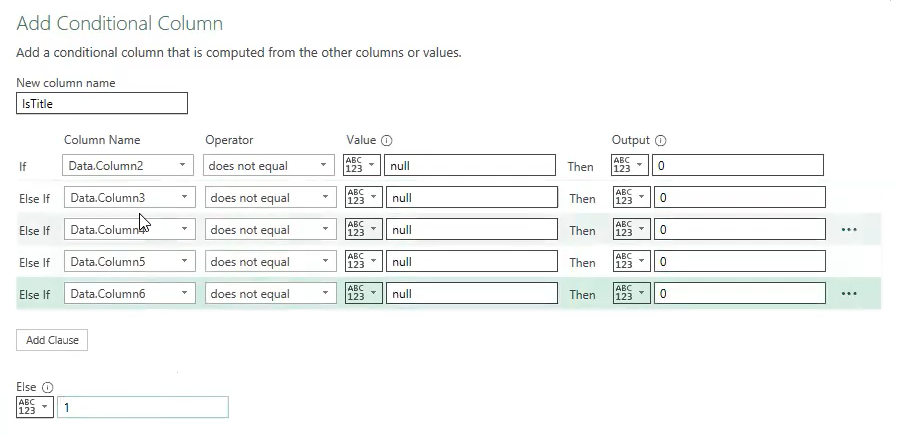


* All on same rows in each sheet (Tab) inside the same worksheet



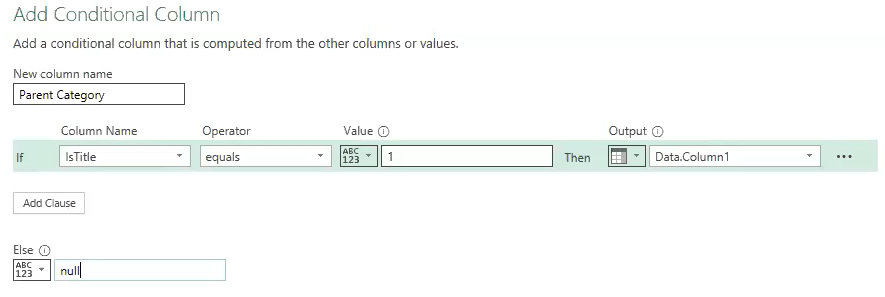
1. File > from workbook > then grab the C0504 – Products
2. Click on the folder icon which will then grab all the sheets. Then transform data
3. We can delete some columns so right click the Data column and select remove other columns
4. Expand the Data column
5. Then go to Home > Remove Rows > Remove Blank Rows
6. Add Column > Conditional Column
7. Call this column IsTitle

* We will create a clause for each column except column1



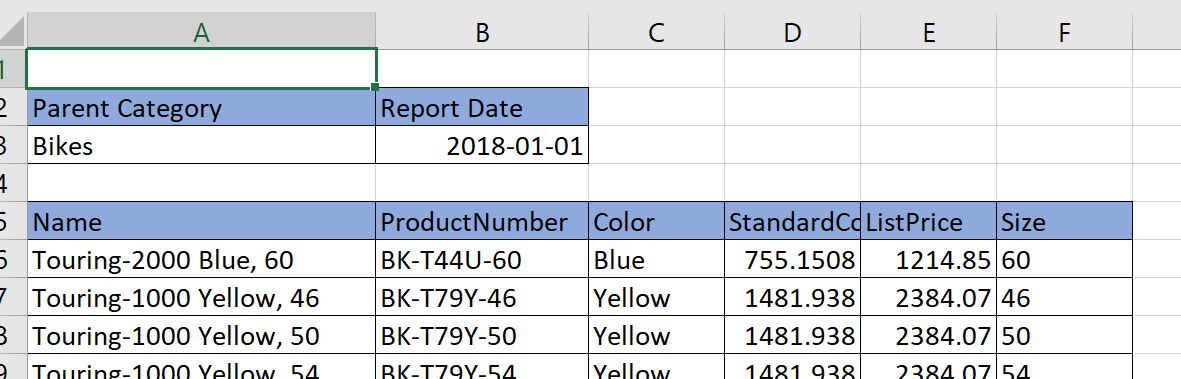
* This will look through the columns and if there are any nulls then it will label it as 1 and everything else as 0. This will indicate it is not a title. Only If all 5 are nulls then it will indicate it is a title. Then click ok

1. Click Add Column > Conditional Column
2. Call it Parent Category
3. Then if IsTitle = 1 then select Column Data.Column1 else make it a null



1. On the Parent Category right click then select Transform > fill > down
2. Now to remove the rows that have the title name and nulls beside them, click the IsTitle and filter and remove the 1
3. Promote first row as headers
4. Remove the IsTitle Column
5. Filter out from the Data.Column1 “Name”
6. Rename the last column to Product Category

## E5-5 – Using Index



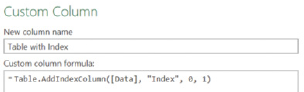
* Same spot on the page on each worksheet

1. Import the file C05E05 – Products
2. Click on the folder to import all the sheets and click transform data
3. Remove all the columns except for Data
4. Then you will only see the Data Column we need to now go to Add Column > Index Column
5. Click that Index Column you created to highlight it. Then in the Formula Bar copy that formula. We will use this soon.
6. Remove the Index Column
7. Now Click Custom Column we will add that formula we copied into the Custom Column Formula

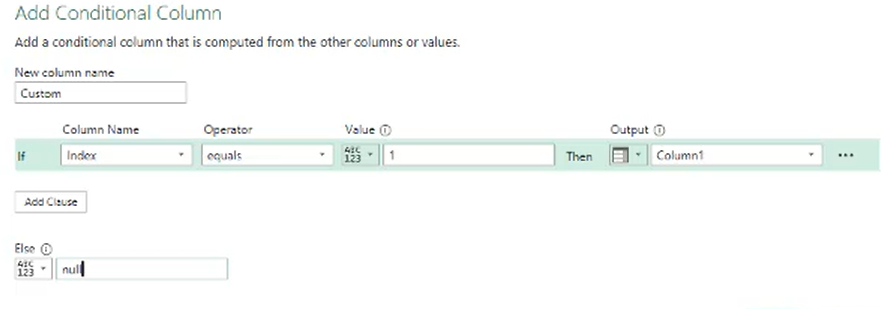


Rewrite the formula to remove removed other columns to [Data]

= Table.AddIndexColumn([Data]), “Index”, 0, 1)

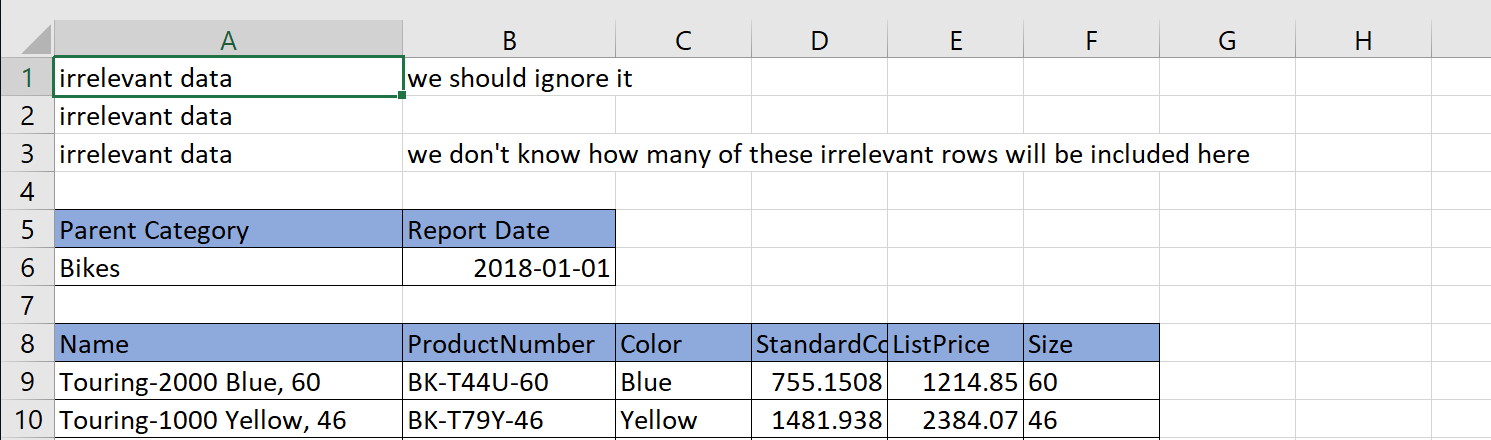


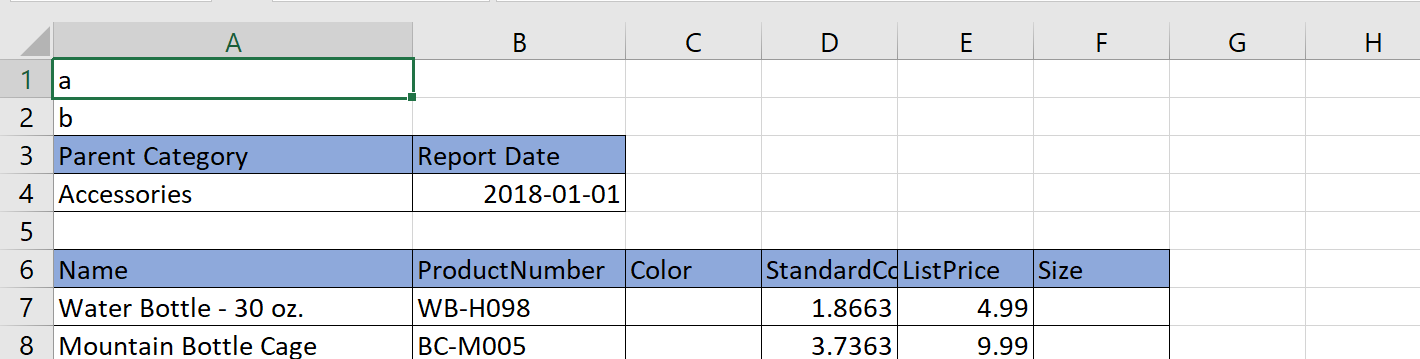
1. In the Applied Steps you can remove the Added Index and Removed Columns
2. You can now remove the Data Column
3. Expand the Custom Column now (do not select/checkmark the use original column name as prefix)
4. Add Column > Conditional Column



1. Right click that Custom Column created then go to Fill > Down
2. Now Click the arrow beside the Index Column and uncheck the index 0, 1 and 2
3. Now Use First Rows As Headers
4. Rename the last column to “Parent Category”
5. Now the second column will say 3 instead of Index but we can filter out the number 3’s because it contains the title of the row
6. Now delete that Index Column

## E5-6 Using Anchor Value





* Files isn’t standardized in different sheets
* Headings start at different points in the sheet but the Parent Category and titles are spelled the same and in the same position

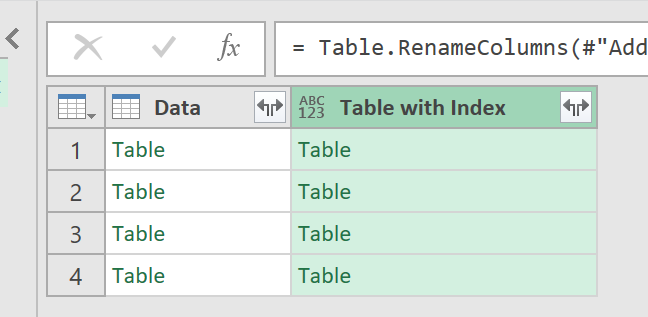
1. Import
   1. file C05E06 – Products
   2. Select the folder
   3. Transform data
2. Add table with Index Column before expanding like be did in E5-5 from steps 1 to 7 in my notes I wrote
   1. Add index Column
   2. Copy the formula then delete that step from Applied Steps

= Table.AddIndexColumn(#"Removed Other Columns", "Index", 0, 1, Int64.Type)

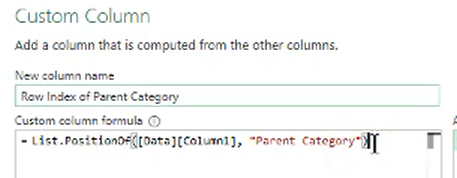
* 1. Then rewrite the code to:

= Table.AddIndexColumn([Data], "Index", 0, 1, Int64.Type)

* 1. Delete the Data Column now

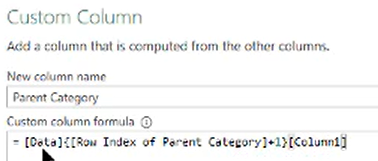


1. Click on the Table with Index then Add Column > Custom Column
2. Call it Row Index of Parent Category (we will be looking for a string of text that matched from the Data Table with a string of text we are looking for called “Parent Category”



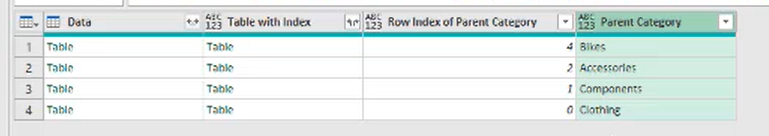
List.PositionOf([Data][Column1], “Parent Category”)

1. We need to do this again Click on Add Column > Custom Column
2. This time we will call it Parent Category

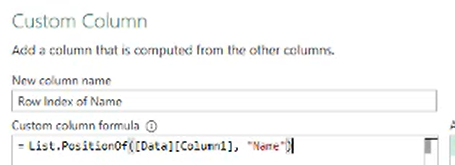


= [Data]{[Row Index of Parent Category] + 1}[Column1]

1. Now you will see all the Parent Category Names Listed



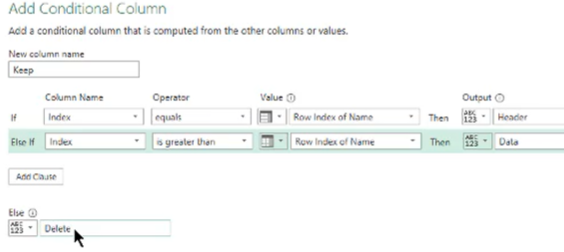
1. We will add another Custom Column for the name header. Add Column > Custom Column – call it Row Index of Name



List.PositionOf([Data][Column1], "Name")

= List.PositionOf([Data][Column1], “Name”)

1. We can delete the Row Index of Parent Category and Data Column
2. Make sure that Data Table is deleted we do not need it anymore.
3. Now we can expand the Table with Index out (uncheckmark Use original column name as prefix)
4. Add a Conditional Column



1. Then click on the Keep column arrow and uncheckmark the Delete
2. Now promote first row as header
3. Now in the Header Column uncheck mark the Header
4. Now we can remove the columns that contain Index, Row Index Indicator, Take the Filter (last row) away
5. Make sure to make the ListPrice and Standard Cost as Currency