WEEK 2 — GET AND TRANSFORM (POWER QUERY) — PART 2

Now that we have imported some data, we can start to clean and organise it to make it easier to analyse later.

KEEP AND REMOVE ROWS

The **Keep Rows** and **Remove Rows** buttons on the **Home** ribbon contain several options for choosing which rows to keep or remove. Often data will have extra rows at the top or a total row at the bottom, or blank rows, that you need to remove before analysis. **Keep Errors** and **Keep Duplicates** can be useful when you are troubleshooting your query.

PROMOTE HEADERS

Usually the first row of your data will contain the column headers. You can **Promote Headers** to get Power Query to recognise these as column headers.

FILTER AND SORT

To remove values from a column, click on the Filter dropdown to the right of the column header. Like in Excel, you can tick/untick specific values or you can use filters based on the data type of the column. You can also sort from this menu. Unlike Excel, sorting by a second column adds to the sort.

SPLIT COLUMN

Sometimes a column will contain multiple pieces of information and you need to extract them into separate columns. The **Split Column** tools can help. You can find them on the **Home** and **Transform** tabs.

If there is a character or characters that separate values, you can use **Split by delimiter**. Newer versions of Power Query allow you to split on lowercase to uppercase, or digit to non-digit.

REPLACE VALUES

You can use **Replace Values** on the **Home** or **Transform** ribbon to replace a value with another. In the **Advanced options** you can choose some special characters. You might need to replace nulls with 0 if you want to perform calculations with your data.

COLUMN FROM EXAMPLES

A powerful addition to recent versions of Power Query is **Column from Examples**. This is like **Flash Fill** in Excel but much smarter. Start typing what you want your new column to look like and Power Query will figure out how to make it from the existing columns. You sometimes need to give several examples to get the right transformation.

ADD CUSTOM AND CONDITIONAL COLUMNS

To create a column based on calculations you can add a **Custom Column**. Here you have all the power of the M language to create exactly what you want. **Conditional Columns** assist you to create columns based on a series of logical tests. You can use a **Custom Column** with **if/then/else** statements if you do not have this feature.

DATA TYPES

When you import some data, Power Query will try to guess what type of variable each column contains. You can also change them manually if you need to by clicking to the left of the column name or on the **Transform** ribbon.

PIVOTING DATA

There are 2 main ways to organise data in spreadsheets: values across the columns (*wide*) and values down the rows (*long*). For example, you might organise sales data with each month having its own column and the departments in the rows. This is a *wide* format. The equivalent *long* format would have columns Department, Month and Sales.

The preferred layout for data analysis is the *long* format but for reports often *wide* is easier to read. In Power Query, going from *long* to *wide* is **Pivoting**. Going from *wide* to *long* is **Unpivoting**. When you **Pivot** you have the choice of how to aggregate the values. **Sum** and **Count** are commonly used, and there is also the **Don't Aggregate** option, which keeps all the values.

Long

Market Type	Month	Total Sales
Farm Stand	January	1994.5
Mobile Market	January	814.15
Farm Stand	February	2606.25
Mobile Market	February	707.9
Farm Stand	March	3076.75
Mobile Market	March	1008.25

Wide

Market Type	January	February	March
Farm Stand	1994.5	2606.25	3076.75
Mobile Market	814.15	707.9	1008.25

TRANSFORM/ADD COLUMN RIBBON TABS

With a bit of exploration, you can see that the **Transform** and **Add Column** ribbon tabs contain some common items. The difference is that the **Transform** tools will modify the current column whereas the **Add Column** tools will create a new column and keep the existing column.