LESSON HANDOUT

Text Analysis in Power BI

Why:

A large amount of data is collected in text form. Text analytics allows you to make use of this data.

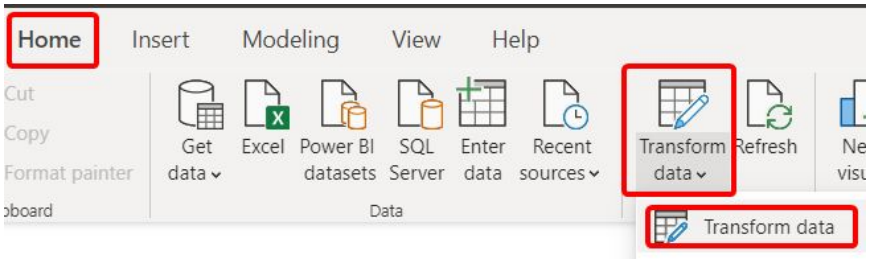
Cleaning and Transforming Text Data in Power Query Editor

Splitting Columns:

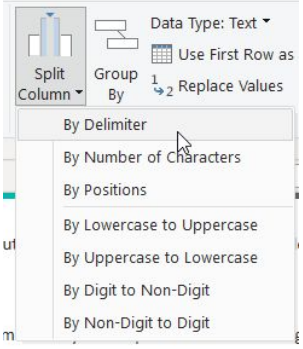
Why: To split text such as City, State, Postcode into separate columns for better data manipulation.

How:

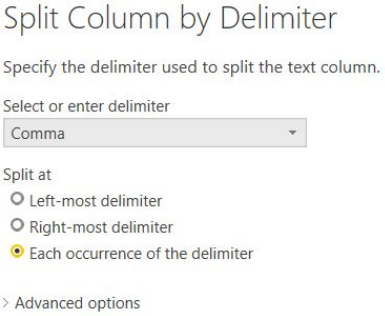
- Open up Power Query Editor - in the ribbon, on the Home tab, under Transform click Transform data > Transform data.



- In Power Query Editor, on the Home Tab, under Transform click 'Split Column' and choose 'By Delimiter'. Note you have other options here, such as 'By Number of Characters' which can be useful.



- Choose the delimiter used to split the column and where the split should occur.

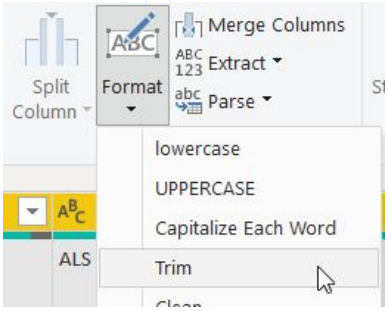


Trim:

Why: Remove unnecessary spaces before and after words for better analysis

How:

- Within Power Query Editor with your column(s) highlighted, in the ribbon navigate to the Transform tab - under Text Column, click Format > Trim.
    - Note: this is similar to the =TRIM formula in Excel.

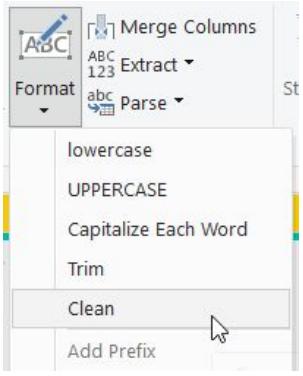


Clean:

Why: Remove unnecessary spaces between paragraphs or lines for better analysis

How:

- Within Power Query Editor with your column(s) highlighted, in the ribbon navigate to the Transform tab - under Text Column, click Format > Clean

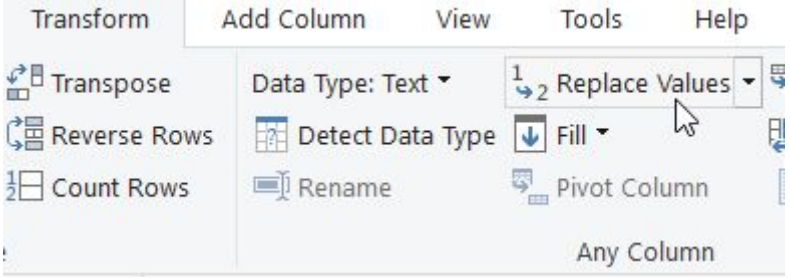


Replace Values:

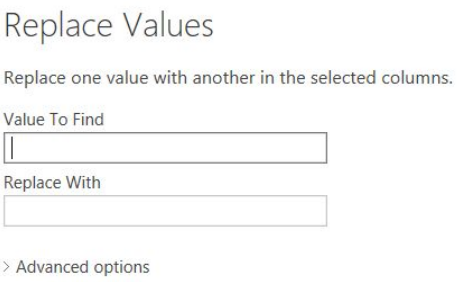
Why: Clean the data by removing unnecessary special character, or replace a specific character or string with another.

How:

- Within Power Query Editor, with your column(s) highlighted, in the ribbon on the Transform tab under Any Column - click 'Replace Values'



- Specify the Value to Find, then the value to Replace With (or leave Blank to replace with nothing, ie removing the Value).

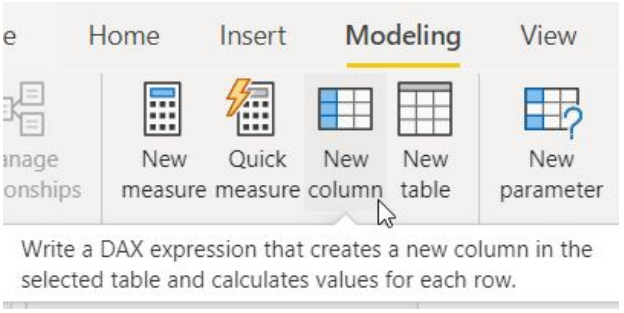


DAX:

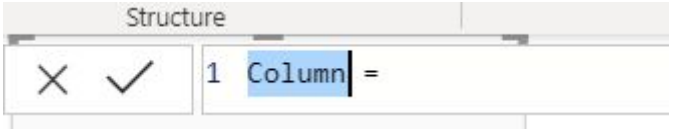
Create a column using DAX:

How:

- On the ribbon, on the Modeling tab, under Calculations, click ‘New Column’.



- Rename your new column and then type in your DAX formula. Some common Formulae and examples are below.



- **DAX:** CONTAINSSTRING(WithinText, FindText)
- **Purpose:** Returns ‘True’ if the specified text is within the string or ‘False’ if it isn’t.
- **Example:** New Column = CONTAINSSTRING(table1(column), “sample text”)
    - Note: When using this function be wary of words within other words, for example, “Excel” will return ‘True’ for not just the word Excel but also for the word ‘excellent’. To fix this, add a space after the word (ie “Excel ”)

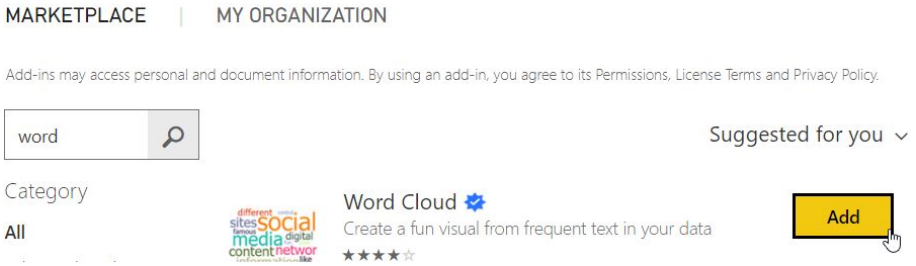
Word Clouds

Create a Word Cloud:

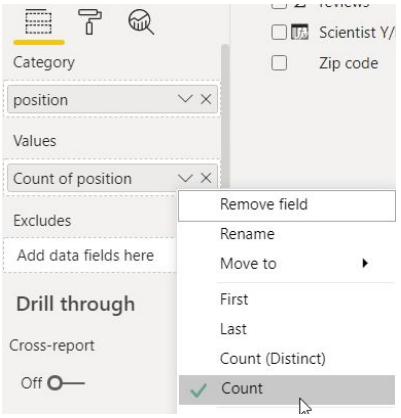
Why: Word clouds allow you to quickly visualise the number of occurrences of a word - the larger the word, the more times it appears in your data.

How:

- Import a word cloud visualisation from AppSource



- Add your text field to the 'Category' section
- (Optional) Add the same text field to 'Values', then update to be a Count



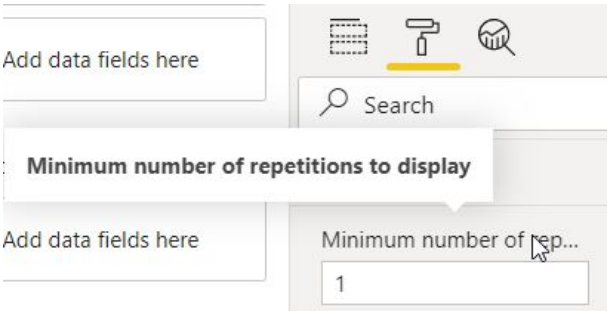
Optimise your Word Cloud with Formatting

Minimum number of repetitions to display:

Why: Useful for filtering out words that only appear a few times, which reduces clutter

How:

- Within the Format tab, under 'General', set the minimum number of repetitions to display

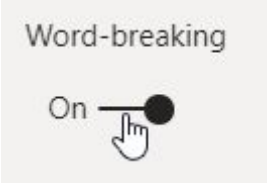


Word-breaking:

Why: Show phrases instead of individual words (such as 'Data Scientist' instead of 'Data' and 'Science' individually)

How:

- Within the Format tab, under 'General', toggle Word-breaking ON/OFF

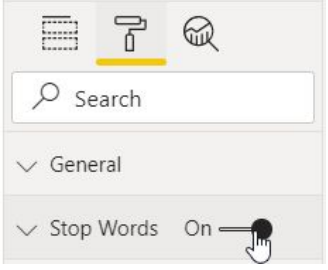


Stop-Words:

Why: Remove words such as 'and', 'the', 'of' etc from your word cloud

How:

- Within the Format tab, under 'Stop Words', toggle ON/OFF



- From here you the option for the Default Stop Words (such as 'and', 'the', 'or' etc) On/Off, and;
- You can also specify words to be filtered out by typing these in the 'Words' section (each word separated by a space). Note: You can clear these filters by clicking on 'Revert to default'.

Stop Words
On
Default Stop Words
On
Words
data experience
Revert to default