# Power BI Notes

Key parts of power bi to master

* Power query editor - for cleaning and preparing the data
* Data modelling – for building the relationships between multiple tables
* Visualisations and filters
* DAX – expressions and formulas for manipulating data in table
* Power bi service

## Power Query Editor

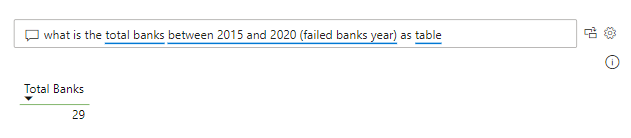
* My step-by-step approach for cleaning data
  1. Ensure excel data is formatted as table before importing into power query editor: Open Excel data -> Highlight table array -> **Home** tab -> **Styles** group -> **Format as Table**
  2. Rename columns with appropriate and understandable names
  3. Ensure appropriate data types are assigned to each column. **NB:** If the date column is displayed in a different country format, changing the data type directly may cause some error values. To convert to your local date format: Revert to original data type without errors -> Click data type icon (next to col name) -> **Using Locale …** -> Change Data Type to Date -> Change Locale to the country in which the date is originally formatted. **NB:** When a dataset has dates, its advisable to extract the month or year into separate columns for trend analysis when needed.
  4. Remove duplicate rows and blank rows: Power Query Editor -> Select all data or specified cols -> **Home** tab -> **Reduce Rows** group -> **Remove Rows** -> **Remove Duplicates** or **Remove** **Blank** **Rows**
  5. Trim values in only text columns to remove white spaces also allow power bi to count empty strings as null values
  6. Check for implausible data (values unlikely to be true. eg. Age: 999 or -14)
     + Review min and max values for numbers: Power Query Editor formula bar -> **Table.Max ("TableName", "Column Name")** eg. **Table.Max (#"Inserted Year", "Year")**. NB: #”Inserted Year” used in the example is a step name. A step name can also be used as a reference to the table since at each step, a change is made to the table and the step represents that table with the new changes. The formula bar can also be used to create a new step. Therefore, entering this block of code to find the min/max value creates a new custom step.
     + To check the sign of col values (+ve/-ve): **Add** **Column** tab -> **From** **Number** group -> **Information** -> Sign. **NB**: +ve value = 1, -ve value = -1, 0 value = 0
  7. Check for invalid data (values entered in wrong format. eg. Name: s@muel)
     + To remove special characters from column: Power Query Editor -> **Add** **Column** tab -> **General** group -> **Custom Column**
     + **Text.Select([fName],{"A".."z","0".."9"})** //maintain alphanumeric chars
     + **Text.Select([fName],{ "0".."9"})** //maintain only nums
  8. Check for missing data in each column using the horizontal bar indicator beneath each header name. Can also be represented as NA, “N/A”, “Not Applicable”
  9. Check for error values in each column using the horizontal bar indicator beneath each header name
  10. Filter unwanted outliers/anomalies
* **Data cleaning processes can be performed in either the Transform tab or Add Column tab. The Transform tab transforms and replaces the original column with the new transformed data. The Add Column** transforms the data into a new column maintaining the original
* **To fix original table headers/column names being used as rows in the imported data:** Power Query Editor -> **Home** tab -> **Transform** group -> **Use First Row as Headers**
* **Transpose** is used when you want your row headers to become column headers. **Transform** tab -> **Table** group -> **Transpose**
* Total number of rows: **Transform** tab -> **Table** group **-> Count Rows**
* **Merge columns**: Power Query Editor -> **Transform** tab -> **Text Column** group -> **Merge Columns**
* **Split column**: Power Query Editor -> **Transform** tab -> **Text Column** group -> **Split Column**
* **Extract text from any position:** -> Power Query Editor -> **Transform** tab -> **Text Column** group -> **Extract**
* **Trim, upper, lower, proper case:** Power Query Editor -> **Transform** tab -> **Text Column** group -> **Format** -> Trim/lowercase/uppercase/capitalize each word
* **Replace values:** Power Query Editor -> **Transform** tab -> **Any Column** group -> **Replace Values**
* **Prefix/Suffix:** Power Query Editor -> **Transform** tab -> **Text Column** group -> **Format** -> Add Prefix/ Add Suffix
* **Compute new column using mathematical operations on other columns:** Power Query Editor -> **Add** **Column** tab -> **General** group -> **Custom Column**
* **Extract Year/Quarter/Month/Day from year:** -> Power Query Editor -> **Add Column** tab -> **From Date & Time** group -> **Date** -> Year/Month/Quarter/Day
* **To display new updates in power query after edit in data source file: Home** tab -> **Query** group -> **Refresh** **Preview** -> **Refresh** **Preview**
* Functions and formulas written in the power query use the M language
* Percentage operations can be applied on number columns in two ways.
  + Option 1: Give me x% value of the column I have selected: Power Query editor -> **Add** **Column** tab -> **From** **Number** group -> **Standard** -> **Percentage**: **(*x is what you put into the field)***
  + Option 2: The col I have selected is the x% value, now give me its 100% value: Power Query editor -> **Add** **Column** tab -> **From** **Number** group -> **Standard** -> **Percentage of**: **(*x is what you put into the field)***
* Modulo function ie. Give me the value of the selected column modulo x: Power Query editor -> **Add** **Column** tab -> **From** **Number** group -> **Standard** -> **Modulo**: **(*x is what you put into the field)***
* **Working with multiple data sources and worksheets**
* **NB:** For any appending, the data involved should always be clean
  + **Append multiple excel/csv files in a folder**: **Get Data** -> **Folder.** 
    - Ensure the folder contains only the excel files or only the csv files you want to append.
    - After selecting the folder containing your files, you can either use **Combine & Transform Data** from the Combine dropdown or **Transform Data** button option to navigate to the next step. However, note that if acol is present in one file and absent from another, **Combine & Transform Data** option automatically removes that extra col which is not present in other files meaning you cannot access the extra col in your power query editor. **Transform Data** option (the ideal option) on the other hand allows you to maintain the extra col and power query editor assigns “null” for rows which do not have any values for the extra col.
    - After clicking the **Transform Data** option, you don’t see the actual table data from the csv/excel files as this data is packed under the **Content** column.
    - To unpack the table data, add a custom column with the formula **Excel.Workbook([Content])** for excel files and **Csv.Document([Content])** for csv files.
  + Append multiple excel sheets in a single workbook/excel file: **Get** **Data** -> **Excel** **Workbook.** 
    - **NB:** The total number of columns can be different across the worksheets
    - The column positionings(index) can be different across the worksheets
    - However, the column name representing a particular information should be equal across the worksheets
* **Column from examples:**
* This means filling the content of a new column with the content retrieved from an existing column. It’s basically doing a flash fill in excel. Now since this creates a new column, then obviously the feature is found under the **Add Column** tab.
* When you input the first example and power query fails to correctly fill out the rest, train the algorithm by inputting a couple more examples under the first one so that it will learn
* **Power Query Applied Steps Pane**
* **Source**: Power query takes the data source file at this step
* **Navigation**: Power query extracts the table from the data source file at this step
* **Changed Type**: Power query assigns the appropriate data types to the table columns at this step. **NB:** Power query is smart to assign the proper types to each column provided all the values in the column meet the condition of the data type. If at this step a column has the ‘**ABC123**’ data type assigned to it, it means not all values in the column met the required data type condition and this helps you to know that there is an invalid data in your column.
* Fill down/up only fills spaces with null values and not empty string spaces
* **M Language**
* You can write M functions inside using Power Query’s advanced editor feature in **Home** tab -> **Query** group -> **Advanced** **Editor** **OR** **View** tab -> **Advanced** group -> **Advanced** **Editor** **OR Add Column** tab -> **General** group -> **Custom Column**
* In the advanced editor, codes under **let** keyword are the numerous steps used to transform the data and the **in** keyword is the name of the step which stores final result
* You can start a whole blank query where you write purely M functions from defining the data source to transforming the data. **Home** tab -> **New Query** group -> **New Source** dropdown -> **Other** **Sources** -> **Blank Query**
* NB: For M functions, all date functions begin with “Date.” prefix. Text functions begin with “Text.” Prefix and Number functions begin with “Number.” Prefix
* <https://learn.microsoft.com/en-us/powerquery-m/power-query-m-function-reference>
* **Drawback of using power query in power bi:** The power query editor previews only the first 1000 rows. It does this to make the editor very responsive when edits are being made. Although the final edits apply to the entire dataset, viewing only 1000 at a time means there could be

some invalid data beyond the 1000 that we couldn’t detect and therefore couldn’t use that to properly clean data

## Visualizations and filters

**Choosing the right chart for visualization**

* Stacked bar chart (Horizontal bar chart) 
* Clustered bar chart (Horizontal bar chart) 
* 100% stacked bar chart (Horizontal bar chart) 
* Stacked column chart (Vertical bar chart) 
* Clustered column chart (Horizontal bar chart) 
* 100% stacked column chart (Horizontal bar chart) 
* Line chart 
* Area chart 
* Stacked area chart 
* Line and stacked column chart 
* Line and clustered column chart 
* Ribbon chart 
* Waterfall chart 
* Funnel 
* Scatter chart 
* Pie chart 
* Donut chart 
* Treemap 
* Map 
* Filled map 
* Gauge 
* Card 
* Multi-row card 
* Kpi 
* Slicer 
* Table 
* Matrix 
* R script visual 
* Python visual 
* Key influencers 
* Decomposition tree 
* **Q&A :** This visual allows you to interact with the data model in your report to ask questions about your data.

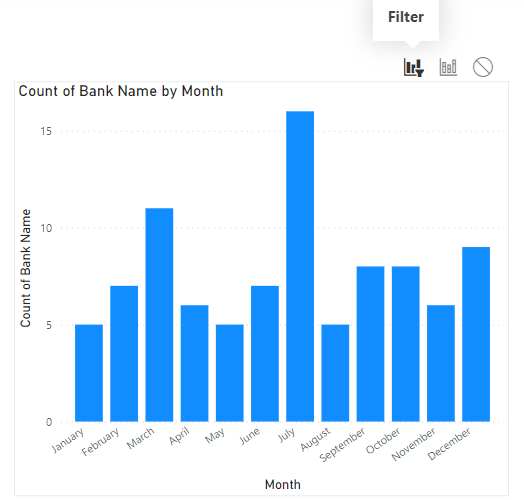
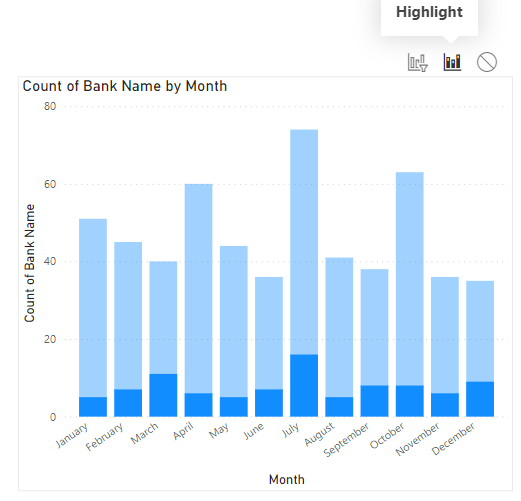
****

Q&A works best with the Synonyms feature which allows to prepare for the case where your users could type in a column or table name different from the actual table or column name but similar meaning. Users often have many terms they use to refer to the same thing, such as total sales, net sales, and total net sales. To add synonyms in Power BI Desktop go to the **Model** view by selecting the **Model** tab and then select a field (column) or table. The **Properties** pane shows the **Synonyms** box, where you can add synonyms. NB: The synonym feature applies specifically to Q&A (and not to Power BI reports in general).

* Smart narrative 
* Metrics (Preview) 
* Paginated report 
* ArcGIS Maps for Power BI 
* Power Apps for Power BI 
* Power Automate for Power BI 
* When building visuals, the metrics are placed inside the ‘values’/ ‘X-axis’ field section of the visualization pane
* Cross highlighting/ cross filtering is where selection of one category in a visual is filtered or highlighted in other visuals. It is enabled automatically in power bi but you can customize this yourself by defining whether you want to highlight the selected category in the other visuals or you want to filter them in the other visuals instead.

1. Click on the visual (A) from which the category will be selected. NB: Not the visual (B) which will display the highlighted or filtered category
2. Go to **Format** menu -> **Edit interactions**
3. At the top right corner of the visual (B), select one of the icons ie. Filter, Highlight or None. PS: In the example below a category was selected in a visual (A). Note the difference between how it highlights or filters that category in visual (B) even though either way displays the same values

Highlight Filter



**Common issues of visualisation**

* If your visual unexpectedly displays the same metric values for all the categories of the data, then it means that category column and metric column do not have a relationship established between them in the data model view

## DAX

* Concatenate joins only 2 columns of the same or different data types into single a text string column.

Eg. DateTime = CONCATENATE([Date],[Time])

To join multiple columns of different data types, use the ampersand (&) operator

Eg. DateTime = [Date] & “ “ & [Time]

Research

* How to append column from one table to another

Additional resources

* <https://xlncad.com/remove-or-extract-special-characters-from-a-data-set-using-power-query/>
* <https://learn.microsoft.com/en-us/power-bi/natural-language/q-and-a-best-practices>
* <https://exceloffthegrid.com/power-query-date-formats/#:~:text=Date%20data%20type,-When%20we%20get&text=Click%20the%20data%20type%20icon,then%20click%20Change%20Type%20%3E%20Date>
* <https://www.spguides.com/power-query-add-column-date/>