Introduction to Excel and Power BI

BOKSS IMHS Internal Training

30-May-2024 @ K83 & Zoom

Chapter 1

Basic Concept of Data Science

Data is everywhere

Type of Data

- Structured Data vs unstructured data
- Quantitative vs. qualitative data

Structured vs unstructured data

Structured data

- Organized
- Tabular format
- Predefined structure
- Text and numbers

Unstructured data

- Unorganized
- No specific format
- No predefined structure
- Text, images, audio, video

Structured vs unstructured data

Structured data

Name	Age	Gender
James	16	Male
Elizabeth	14	Female
Thomas	17	Male

Unstructured data

There are three students named James, Elizabeth, and Thomas. Their respective genders and ages are male 16, female 14, and male 17.

Quantitative vs. qualitative data

Quantitative data

- Numberical data
- Count, measure, percentage

Examples

- Age
- Temperature

Qualitative data

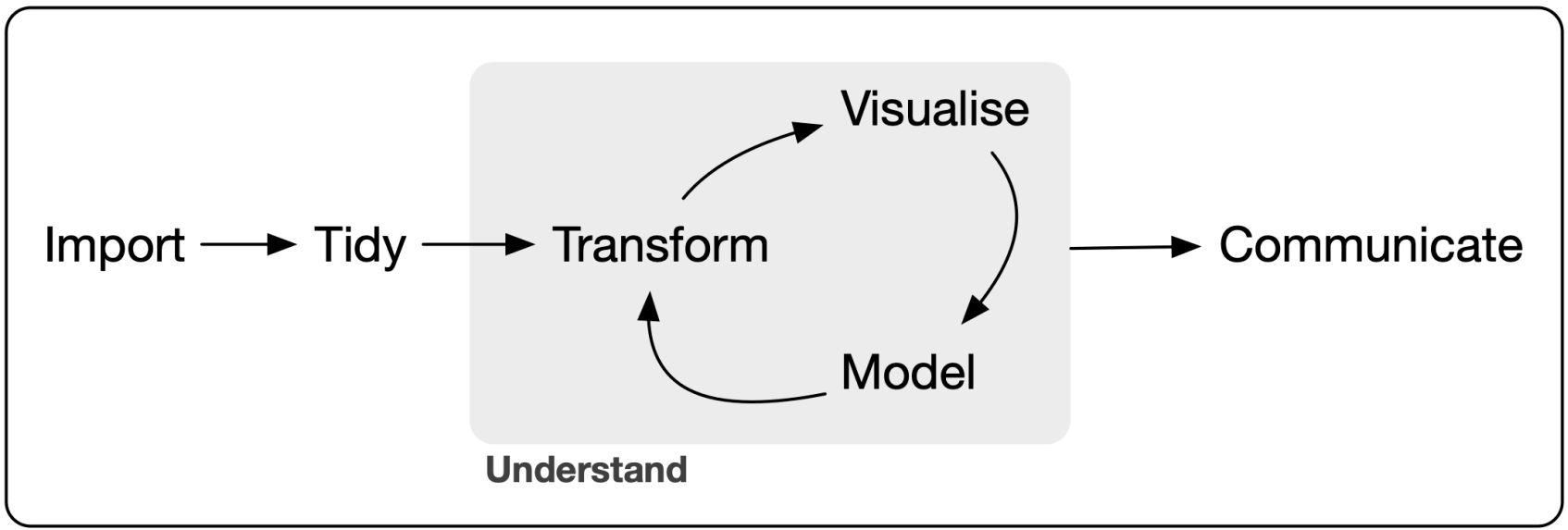
- Categorical data
- Group into categories

Examples

- Gender
- Holiday

Workflow of Data Science

"R for Data Science" (Wickham and Grolemund 2017)



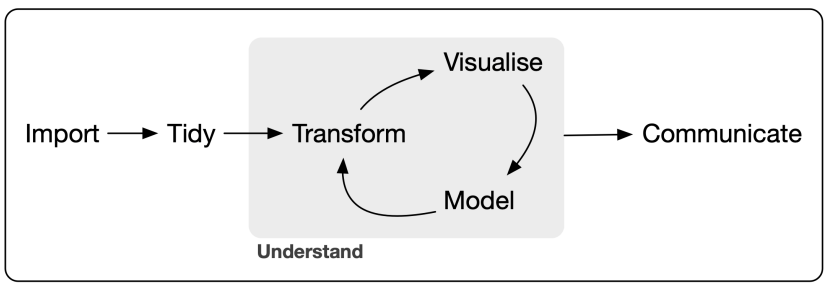
Workflow of Data Science

Import

- Open the file using one software
  - .csv / .tsv / .txt
  - .xlsx / .xls
  - .json

Tidy

- Format in tabular
  - Every column is a variable
  - Every row is an observation
  - Every cell has a single value



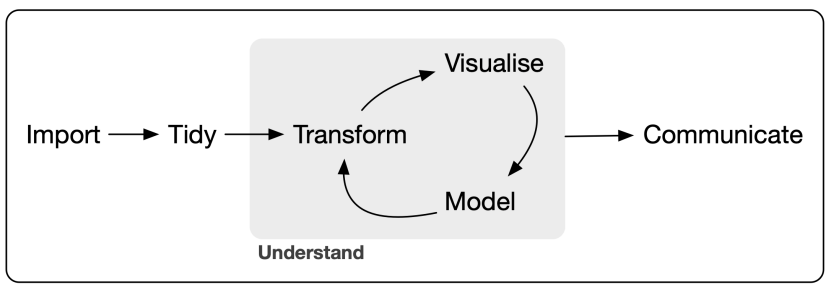
Workflow of Data Science

Transform

- Create new column
- Re-group categories
- Filter your sample
- Combine tables
- Handle missing data

Visualize

- Get the insight of your data
- Check the data quality
  - Missing (Completeness)
  - Validity / Accuracy



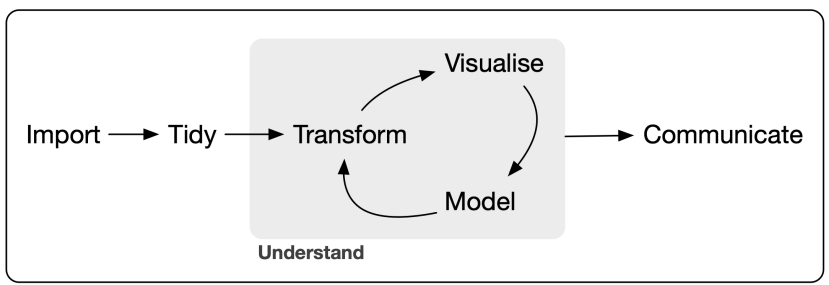
Workflow of Data Science

Modeling

- Descriptive (Business Intelligence)
- Out-of-scope
  - Predictive (Machine Learning)
  - Interference (Statistic)

Communicate

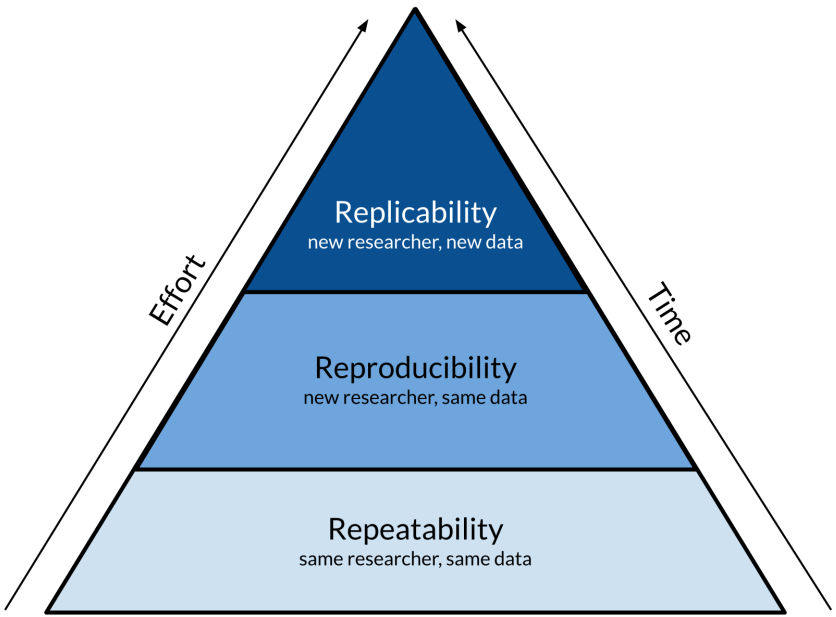
- Making (interactive) dashboard
  - Excel, Google sheet
  - Power BI, Looker, Tableau
  - R, Python



Common data cleaning challenges

- Missing data
- Inconsistent
  - units
  - format
  - spelling variations
  - typo
- Outliers and errors
- Duplicate
- Unstructured
- Wrong data types

3-R's in data science



Based off of a figure from Essawy et al, 2020 https://doi.org/10.1016/j.envsoft.2020.104753

Chapter 2

Some Basic Tricks of Excel

Edit, Copy and Paste

Method 1

1. Select the cell
2. Make your edit, copy, or paste
3. When you are editing, **arrow** keyboard will select other cell

Method 2

1. Double click (F2) the cell
2. Make your edit, copy, or paste
3. When you are editing, **arrow** keyboard will be as usual

Remark

- You can only use the same method to copy and paste

Referencing

Absolute referencing

- **constant** when copied to other cell

```
1   =$A$2 / =A$2 / =$A2
```

Relative referencing

- **change** when copied to other cell

```
1   =A2 / =Sheet1!A2
```

Other topic

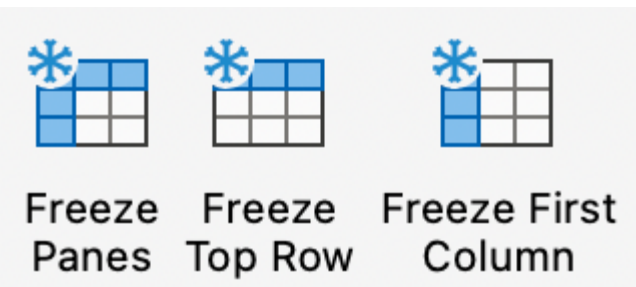
- Multiple lines in one cell
- Freeze cell
- Custom format
- F4 Hotkey
- Resize width and height
- Paste special

Multiple lines in one cell

- Press Alt + Enter

Freeze cell

- View ->
  - Freeze Panes (Custom the Freezing behavior)
  - Freeze Top Row
  - Freeze First Column

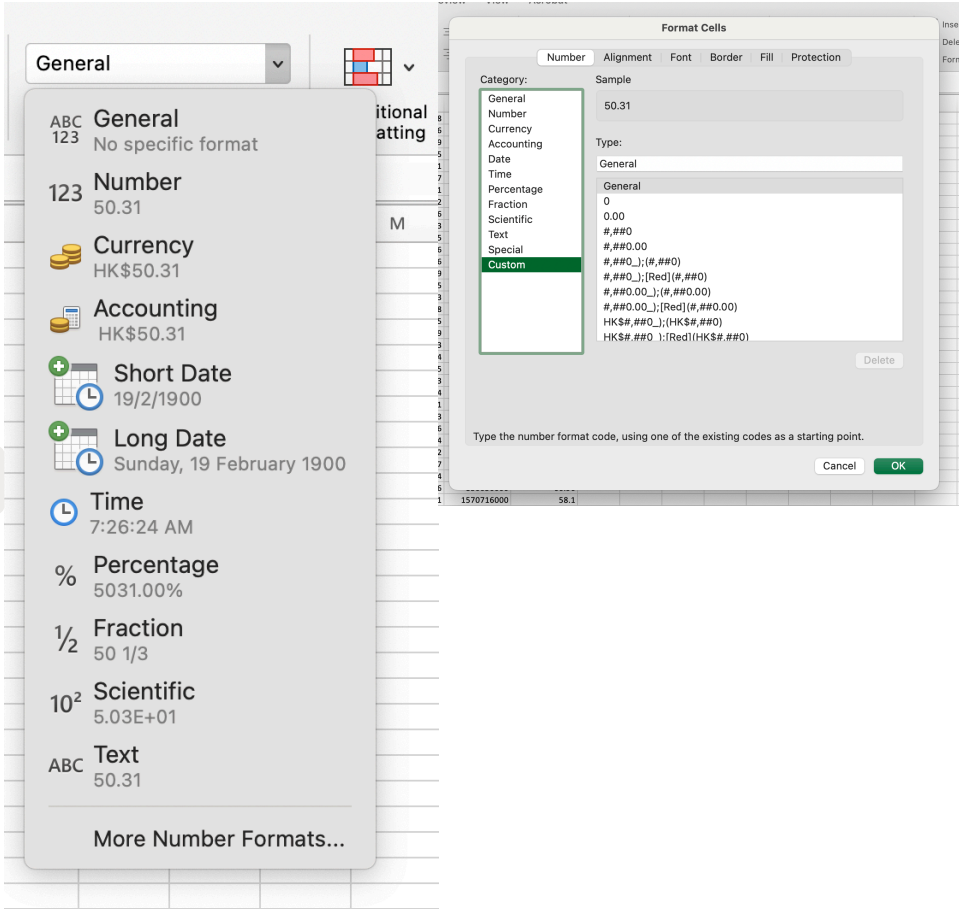


Formatting

- Build-in format
- Custom format
- Using formula

```
1   =TEXT(A2, "[$-404]aaaa")
```

reference-date-format



F4 Hotkey

- Make previous action

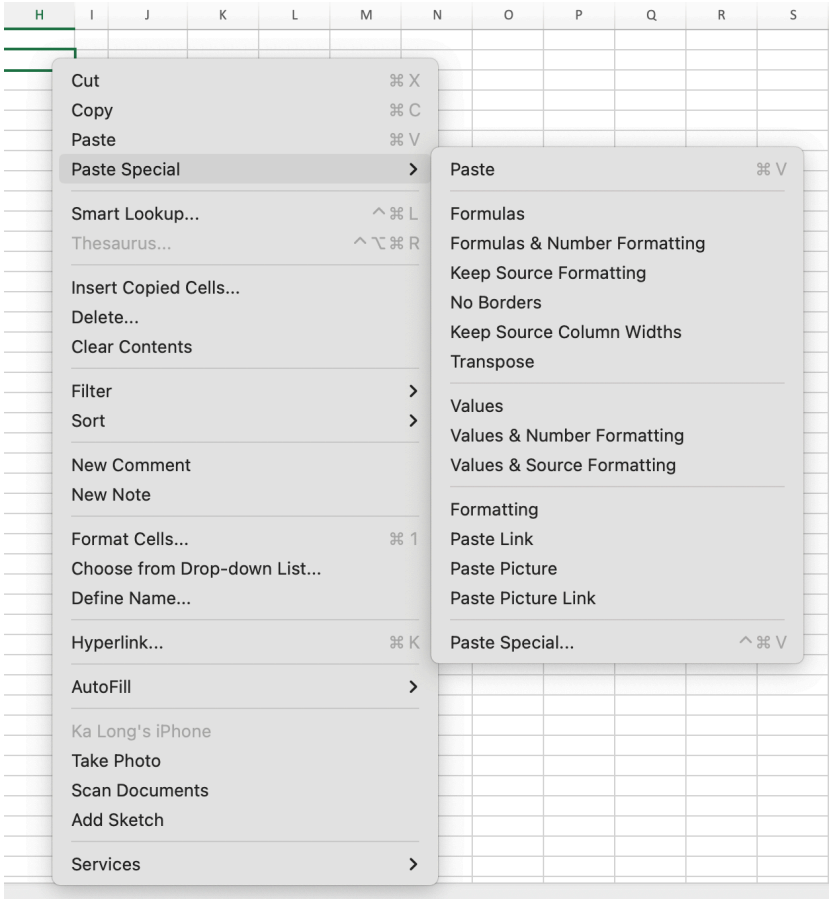
- Change referencing style

Resize width and height

1. Select all
2. Double click the height line or width line

Paste special

- Transpose
- Values



Chapter 3

Using Table in Excel

Feature of Table in Excel

- Column header row
  - Unique header name
- Body
- Total row
  - Disable by default
  - Build-in function (SUM / AVERAGE)
  - Custom function

	A	B	C	D	E
1	OrderNo	SalesOrderLineKey	OrderQuantity	ItemCost	ItemPrice
2	SO43697	43697001	1	$2,171.29	$3,578.27
3	SO43698	43698001	1	$1,912.15	$3,399.99
4	SO43699	43699001	1	$1,912.15	$3,399.99
5	SO43700	43700001	1	$413.15	$699.10
6	SO43701	43701001	1	$1,912.15	$3,399.99
7	SO43702	43702001	1	$2,171.29	$3,578.27
8	SO43703	43703001	1	$2,171.29	$3,578.27
9	SO43704	43704001	1	$1,898.09	$3,374.99
10	SO43705	43705001	1	$1,912.15	$3,399.99
11	SO43706	43706001	1	$2,171.29	$3,578.27
12	SO43707	43707001	1	$2,171.29	$3,578.27
13	**Total**		**11**	**$20,816.29**	**$35,565.40**

Benefit of using Table

Structured referencing

- Automatically updates as data is added

```
=[@colName]
```

Data quality

- Automatically add new column
- Build-in filtering and sorting
- Data validation automatically updates
- Re-size table to add or remove data

Chapter 4

Basic Functions in Excel

Type of functions in Excel

- Operations
- Command button in Ribbon
  - Data validation
  - Conditional formatting
  - Remove duplicates
  - Flash fill
  - Split text to columns
- Function
  - Aggregate
  - Text
  - Date & Time
  - Maths
  - Logical
  - Lookup

Operations

Symbol	Operation
+	Addition
-	Subtraction
*	Multiplication
/	Division
^	Exponentiation

Symbol	Operation
>	Greater than
<	Less than
>=	Greater than or equal to
<=	Less than or equal to
<>	Not equal
=	Equal to

Aggregate function

```
=SUM(A2:A10)
=AVERAGE(A2:A10)
=MAX(A2:A10)
=MIN(A2:A10)
=COUNT(A2:A10)
=COUNTBLANK(A2:A10)
```

Text function

```
=LEFT()
=RIGHT()
=TRIM()
=CLEAN()
=CONCAT()
=CONCATENATE()
=TEXTJOIN()
=TEXTSPLIT()
=UPPER()
=LOWER()
=PROPER()
=LEN()
=REPLACE()
=SUBSTITUTE()
```

Date & Time function

Date

```
=DATE()
=YEAR()
=MONTH()
=DAY()
=DAYS()
=TODAY()
=WEEKDAY()
=WEEKNUM()
```

Time

```
=TIME()
=NOW()
=HOUR()
=MINUTE()
=SECOND()
```

Maths function

```
=ROUND()
=ABS()
=INT()
```

Logical function

```
=AND()
=OR()
=NOT()
=ISNUMBER()
=ISERROR()
=ISERR()
=ISBLANK()
=IF()
=IFS()
=IFERROR()
```

Logical + Aggregate function

[Aggregate][Logical]()

```
=COUNTIF()
=COUNTIFS()
=SUMIF()
=SUMIFS()
=AVERAGEIF()
=AVERAGEIFS()
```

D[Aggregate]()

```
=DMAX()
=DMIN()
```

Lookup function

```
=VLOOKUP()
```

Other useful function

```
=INDEX()
=INDIRECT()
=OFFSET()
=LARGE()
=SMALL()
=ROW()
=ROWS()
=COLUMN()
=COLUMNS()
=CHOOSE()
=SEARCH()
=FIND()
=MATCH()
```

Other useful function only in Excel 2021 or Excel Web

```
=SORT()
=SORTBY()
=UNIQUE()
=FILTER()
=XLOOKUP()
=XMATCH()
=SWITCH()
```

Wildcards in Excel

Use to match pattern in function

- Find and Replace
- Conditional Formatting
- Filter
- SEARCH()
- XMATCH()
- [Aggregate][Logical]()

Symbol	Meaning
?	1 character
*	0 or more character
~	Escape whilcards

Use case

- Find the unique value
- Combine two table (Vlook + Column)
- Data validation with function
- Conditional formatting with function

Pivot Tables

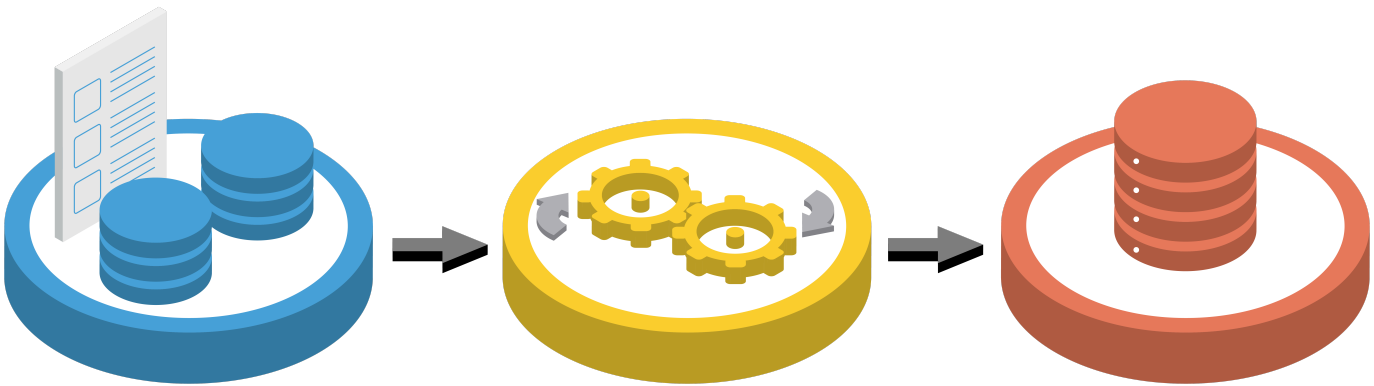
Benefit

- Aggregate and organize data in dynamic tables
- Transform rows to columns, or vice versa
- Group, filter, aggregate without need to make changes

Chapter 5

Power Query in Excel

Big data era

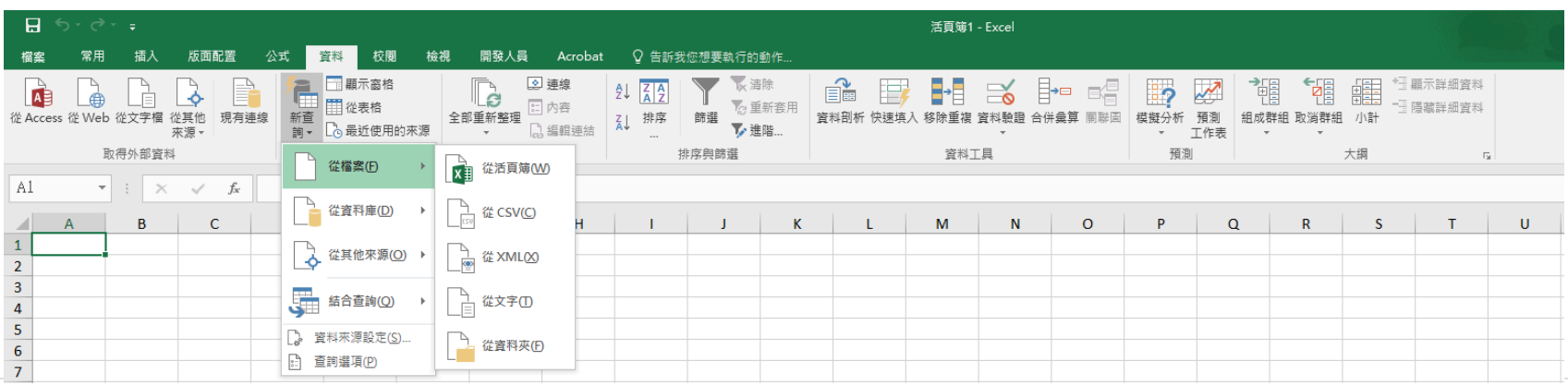


Benefit of power query

- Combine data from different source
- Keep track of every step in the workflow
- Easy to do every step in workflow
- Update the analytics when the data source update
- Ensure the reproducibility

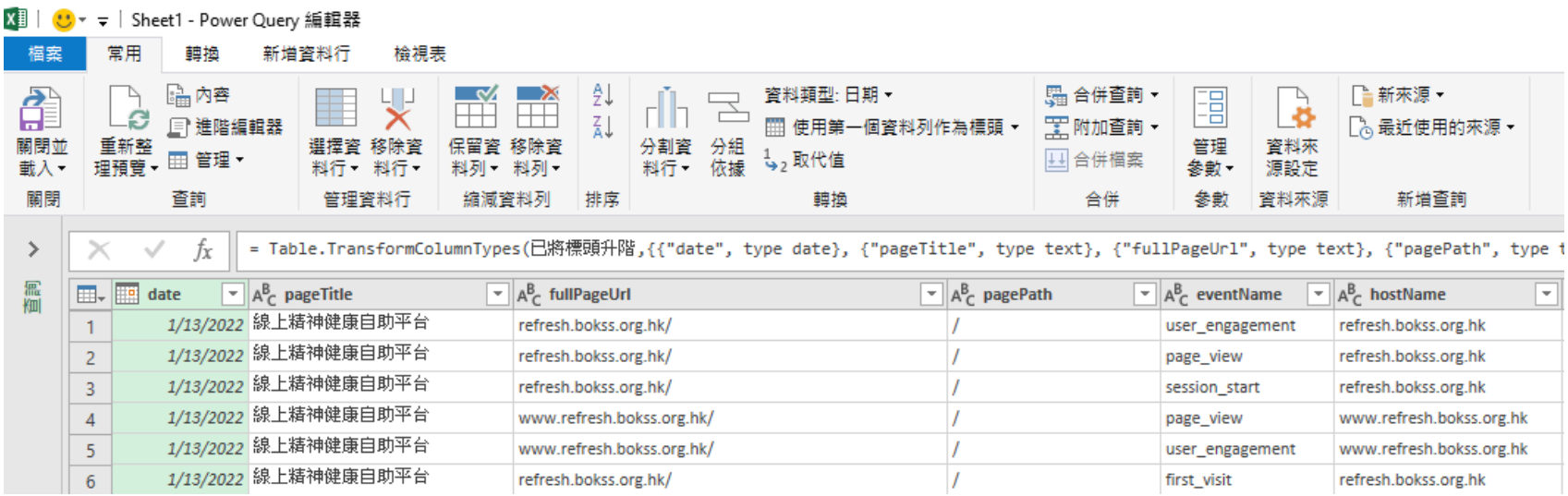
Import data

- Single file
- Multiple files with same structure
- Multiple sheets in Excel with same structure
- File from Onedrive Business (BOKSS)



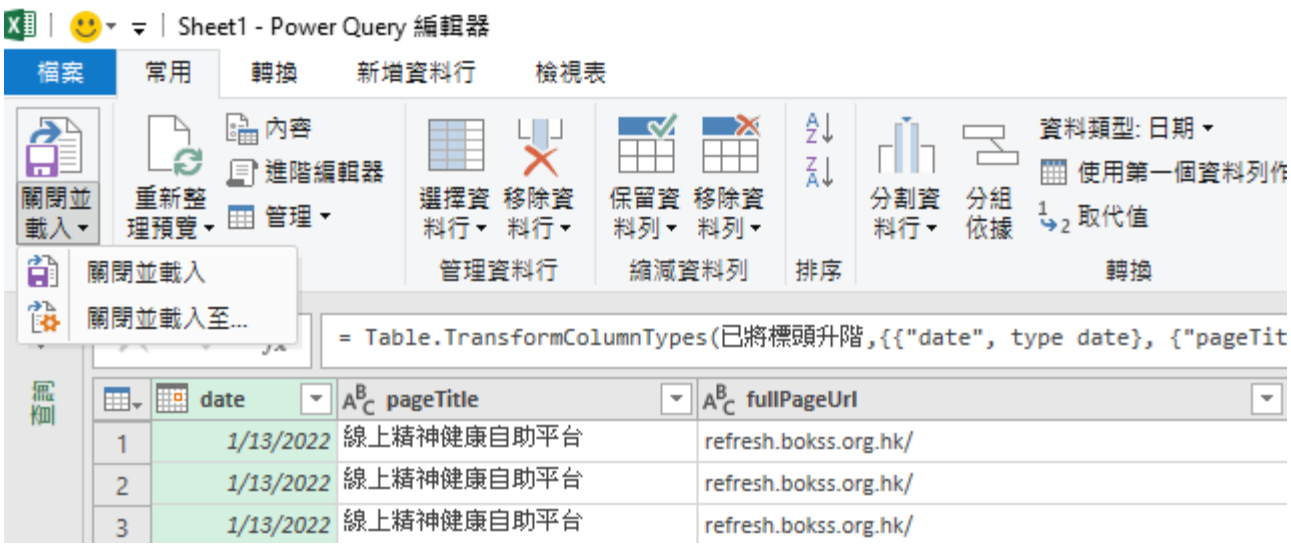
Transform

- Modify current column
- Add new column



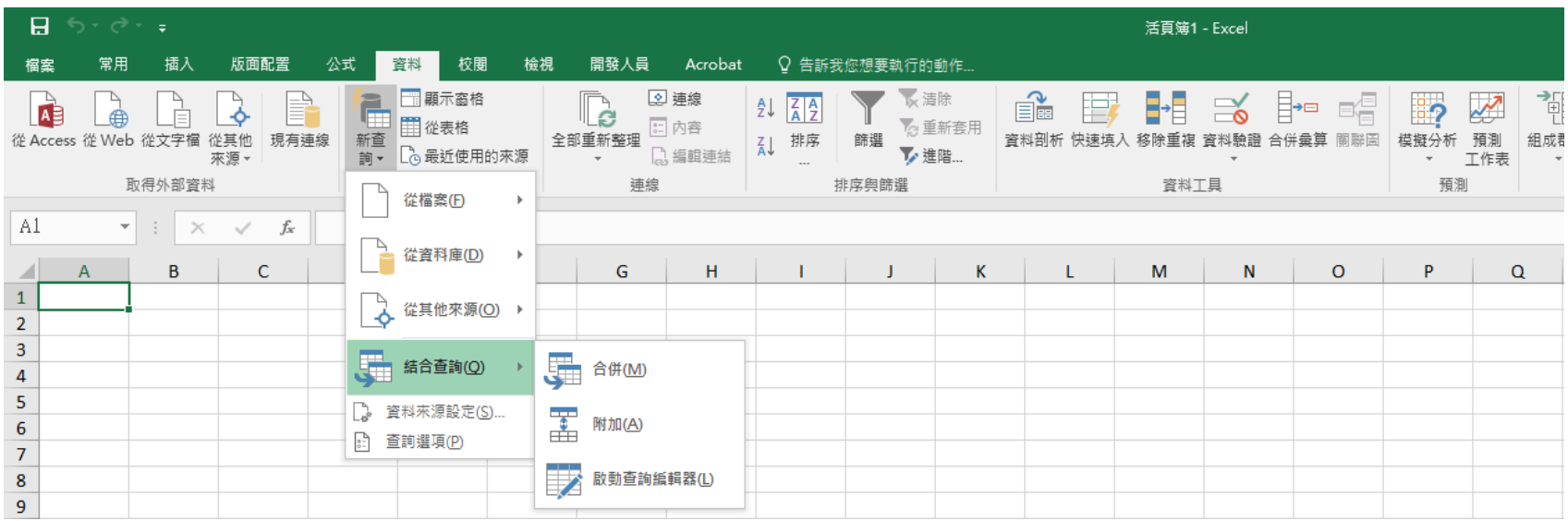
Load

- Load to excel sheet
- Load to excel connection



Combine query

- Merge (Join or map)
- Append



Different type of join

- Full join
- Left join or right join
- Inner join
- Anti join

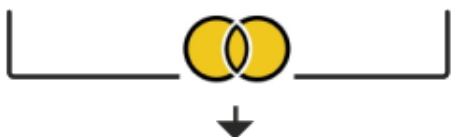
Full join

Left Table

Date	CountryID	Units
1/1/2020	1	40
1/2/2020	1	25
1/3/2020	3	30
1/4/2020	2	35

Right Table

ID	Country
1	USA
2	Canada
3	Panama
4	Spain



Merged Table

Date	CountryID	Units	Country
1/1/2020	1	40	USA
1/2/2020	1	25	USA
1/4/2020	2	35	Canada
1/3/2020	3	30	Panama
null	*null*	*null*	Spain

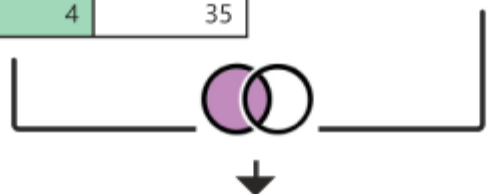
Left join (Right join)

Left Table

Date	CountryID	Units
1/1/2020	1	40
1/2/2020	1	25
1/3/2020	3	30
1/4/2020	4	35

Right Table

ID	Country
1	USA
2	Canada
3	Panama



Merged Table

Date	CountryID	Units	Country
1/1/2020	1	40	USA
1/2/2020	1	25	USA
1/3/2020	3	30	Panama
1/4/2020	4	35	*null*

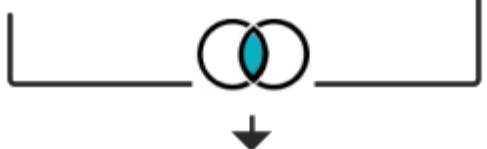
Inner join

Left Table

Date	CountryID	Units
1/1/2020	1	40
1/2/2020	1	25
1/3/2020	3	30
1/4/2020	2	35

Right Table

ID	Country
3	Panama
4	Spain



Merged Table

Date	CountryID	Units	Country
1/3/2020	3	30	Panama

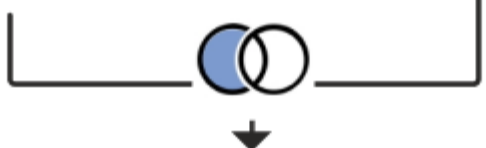
Anti join

Left Table

Date	CountryID	Units
1/1/2020	1	40
1/2/2020	1	25
1/3/2020	3	30
1/4/2020	2	35

Right Table

ID	Country
3	Panama
4	Spain



Merged Table

Date	CountryID	Units	Country
1/1/2020	1	40	*null*
1/2/2020	1	25	*null*
1/4/2020	2	35	*null*

Make your decision

Chapter 6

Power BI Navigation

Report View

1. Canvas
2. Filters Pane
3. Visualization Pane
    - Build visual
    - Format page
4. Data Pane
5. Page Overview

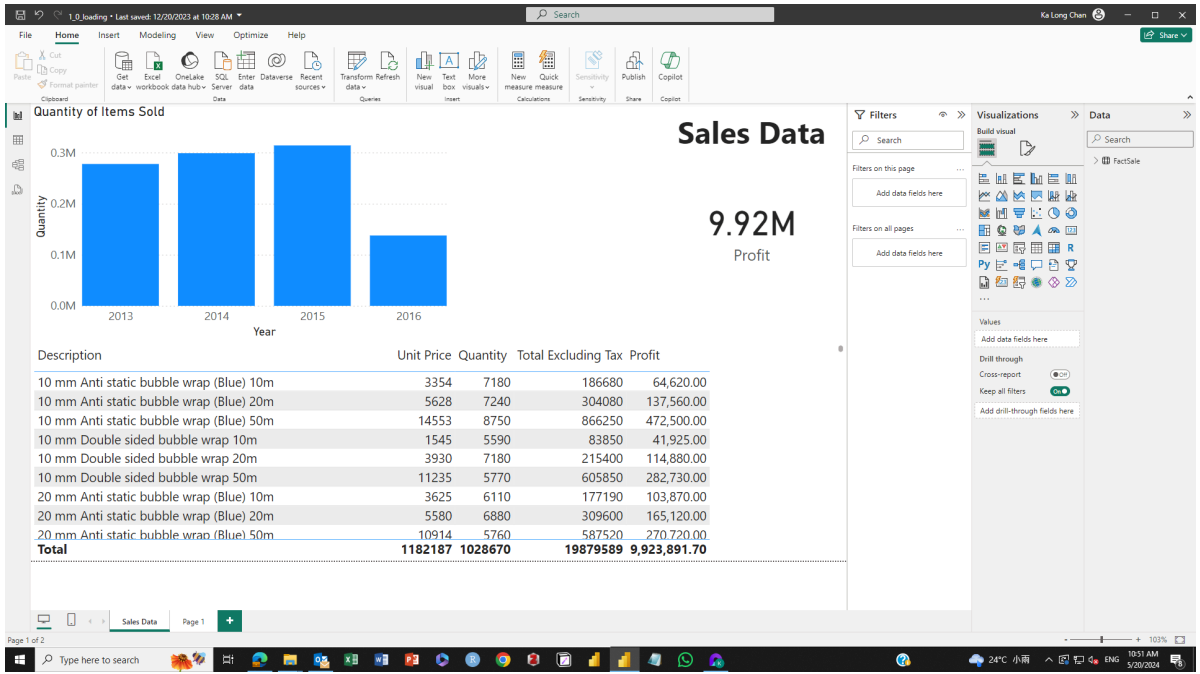
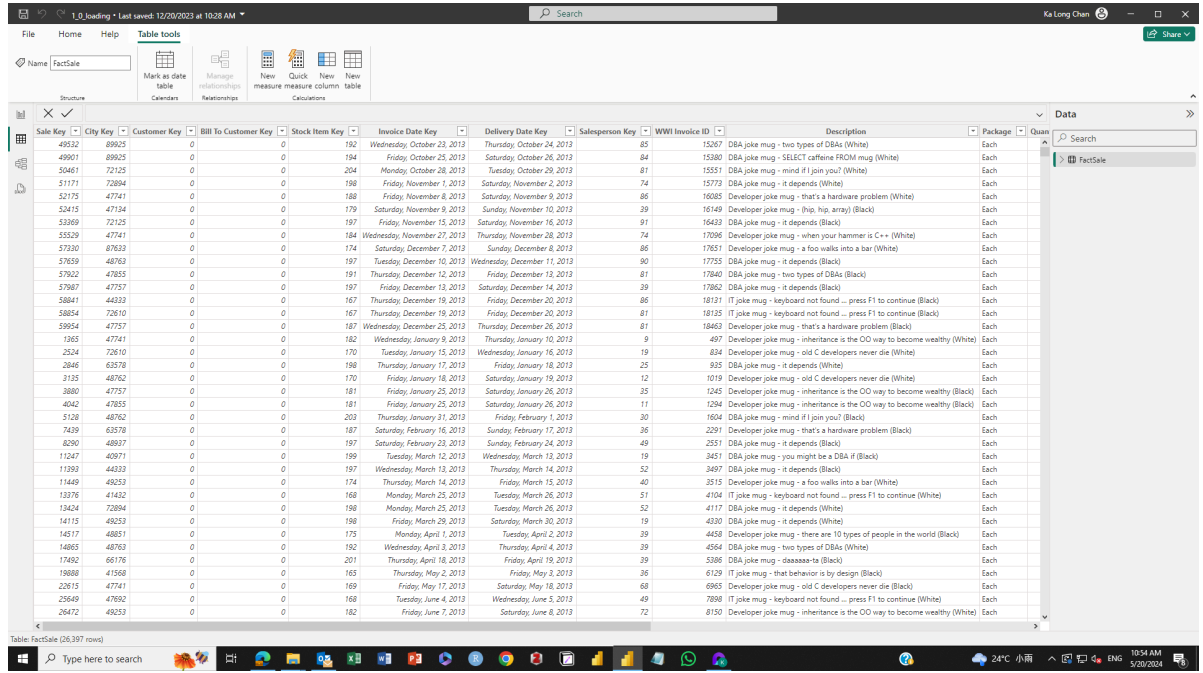


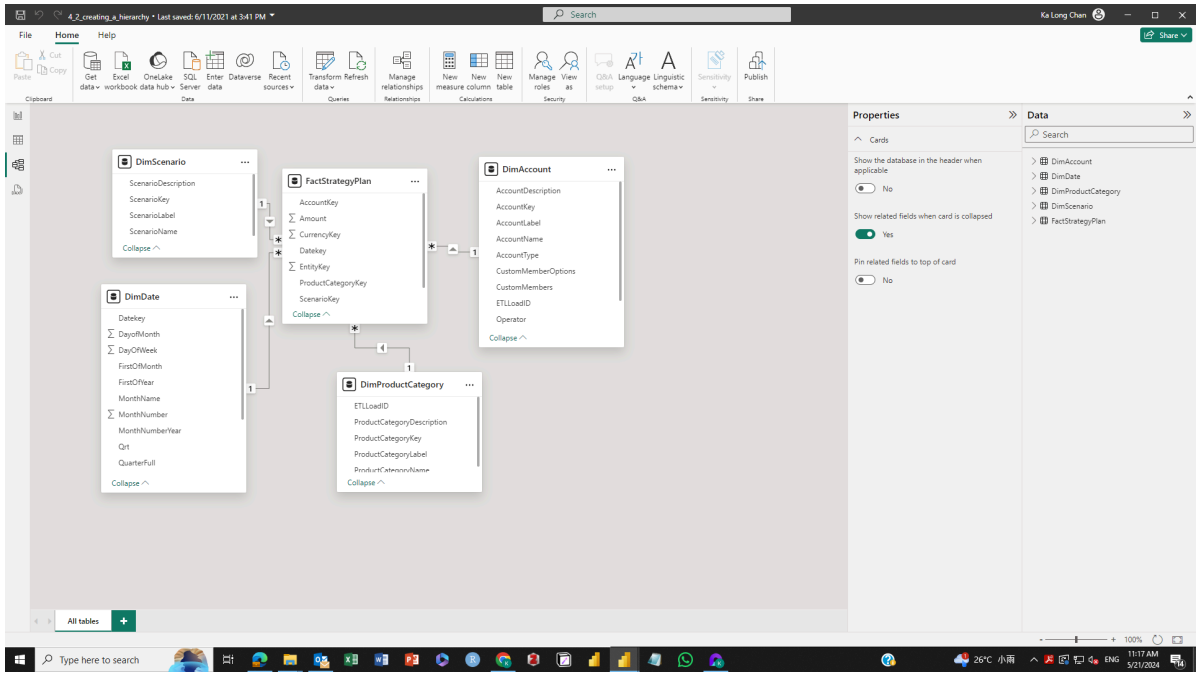
Table View

1. Data Grid
2. Data Pane
3. DAX Formula bar
4. Table Tools

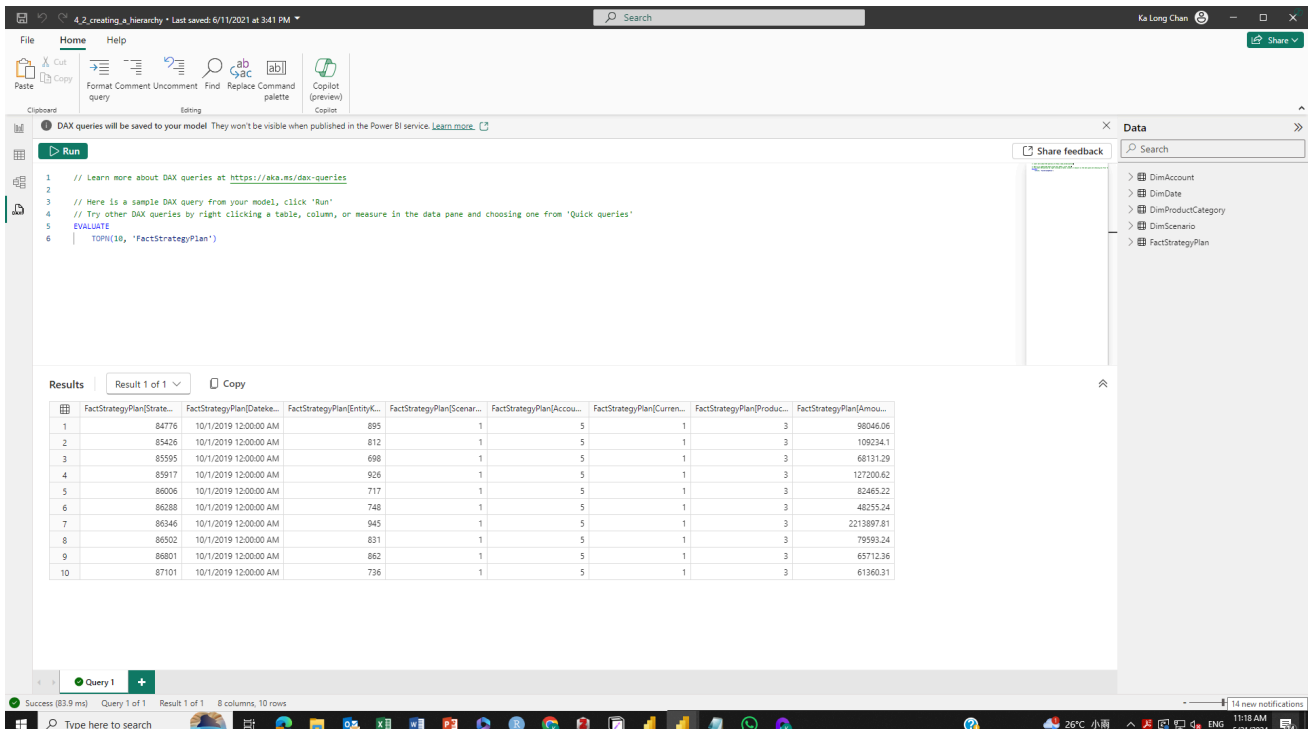


Model View

1. Data Model
2. Data Pane
3. Properties Pane

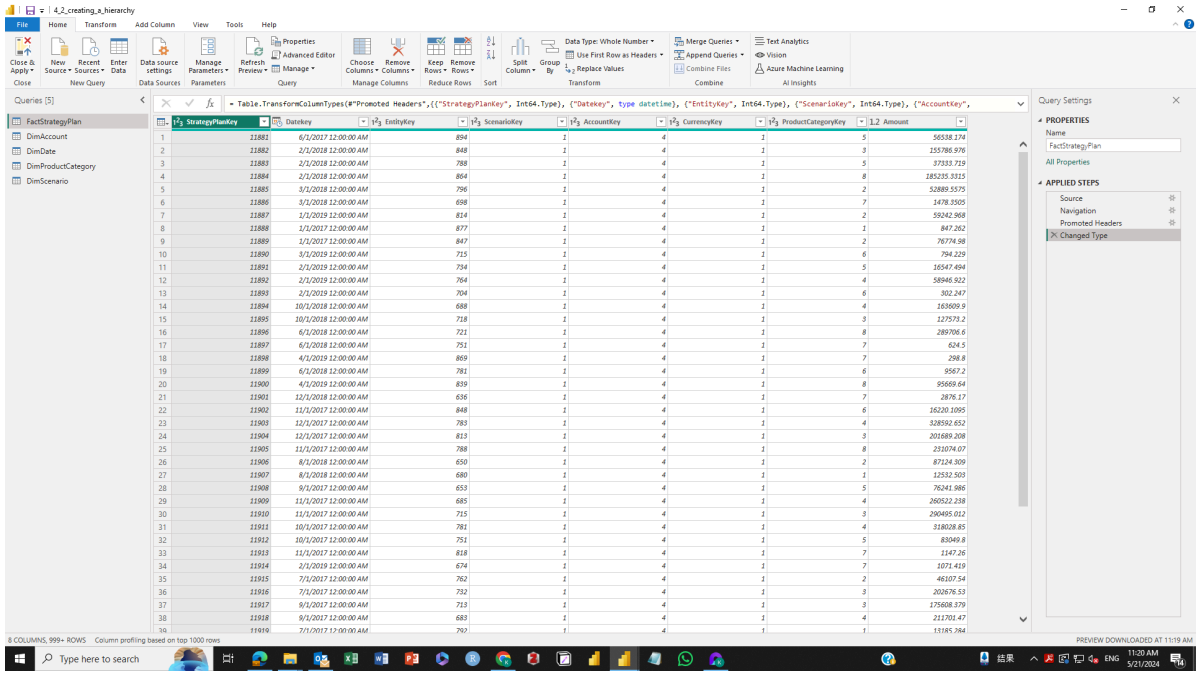


DAX Query View



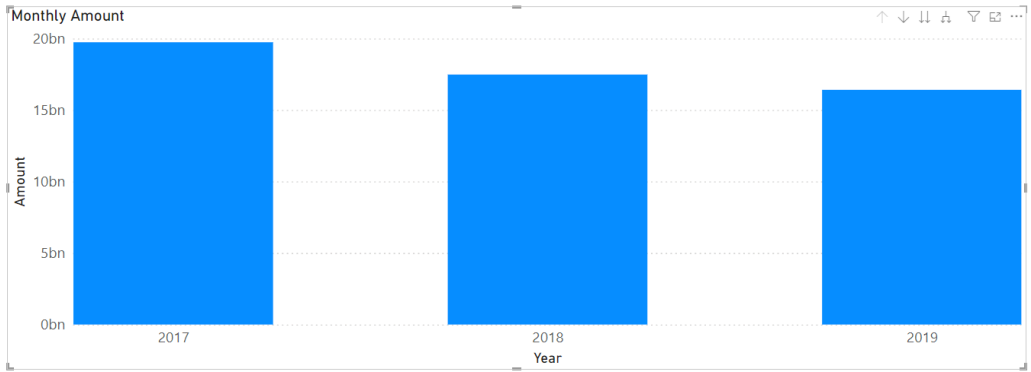
Power Query Editor

1. Query Editor Ribbon
2. Queries Pane
3. Query Setting
4. Status Bar
5. Table Pane



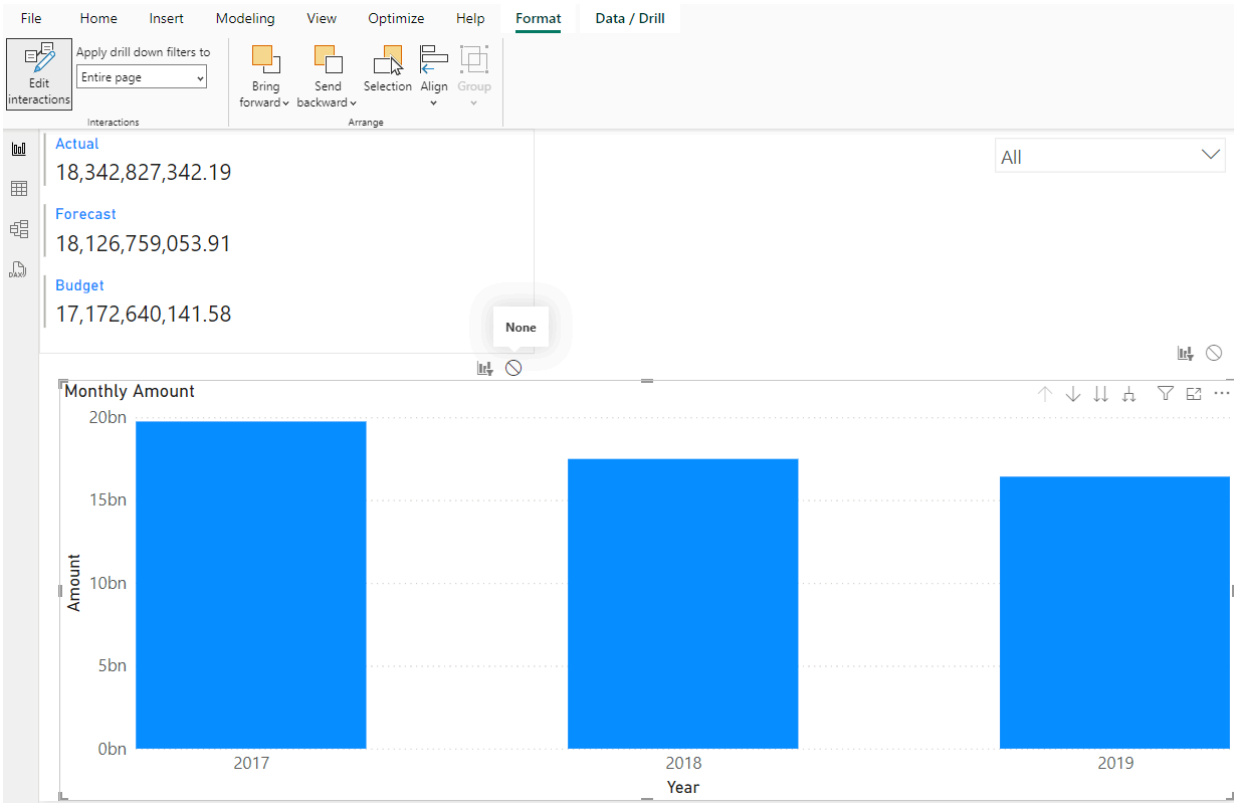
Drilling Down

1. Date
    - Auto create Date Table
    - Mark as Date Table
    - Using DAX
    - Using Power Query (M)
2. Custom Hierarchies
    - Department -> Unit -> Team



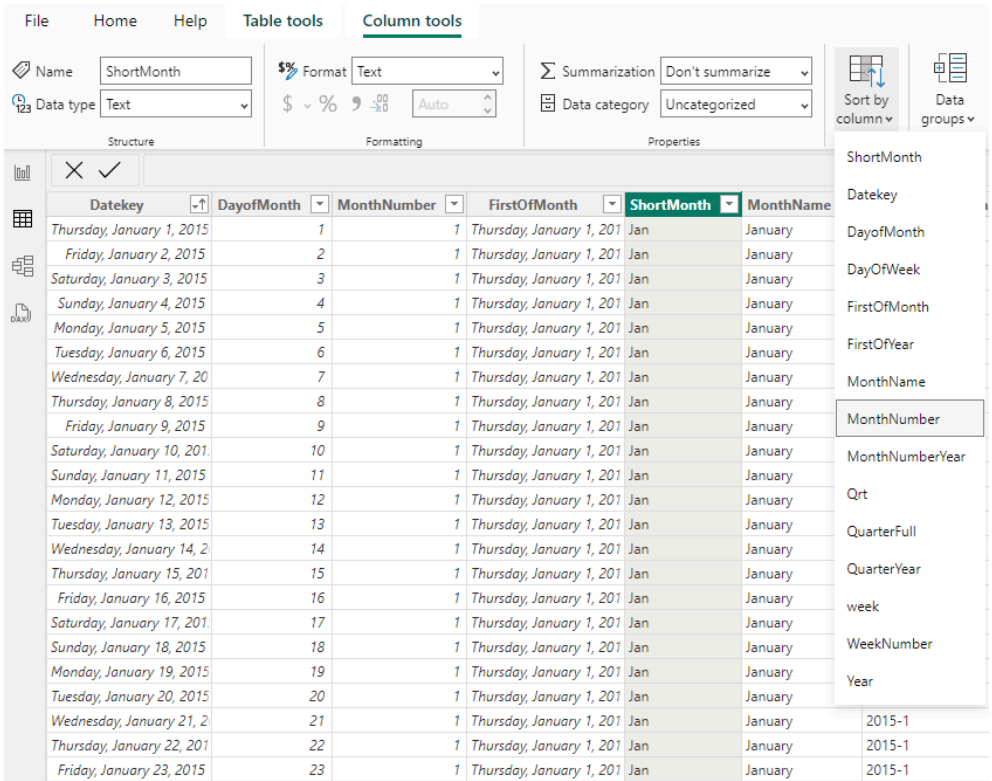
Filter

1. Visual Level
2. Page Level
3. Report Level
4. Slicer
5. Interactions



Sorting

1. Sort in Table View
2. sort in Visual
3. Sort by column



Chapter 7

DAX in Power BI

What is DAX

- **D**ata **A**nalysis e**X**pressions
- Based on Excel formulas and functions
- Create calculated columns, measures, tables
- DAX reference
  - https://docs.microsoft.com/en-us/dax/dax-function-reference

Calculated columns

- Calculate at row level
- Add new column to existing table
- Based on Excel formulas and functions
- Calculate when data is refreshed

```
new_column = Sales[Price] + Sales[Tax]
```

Calculated measures

- Aggregates multiple rows
- Add a new field for visualization
- Calculated when you interact with the visuals
- Two ways
  - from scratch
  - use Quick Measure

Context of DAX

Aspect	Custom column	Calculate column	Calculated table	Measure	Visual calculation
Language	M	DAX	DAX	DAX	DAX
Computed at	Data refresh	Data refresh	Data refresh	On demand	On demand
Persistence	Results saved	Results saved	Results saved	Calculated as required	Calculated as required
Context	Row	Row	Row	Filter	Visual
Stored in	Table	Model	Model	Model	Visual
Changes with user interaction in report	No	No	No	Yes	Yes
Usage	Slicers, filters, rows, columns	Slicers, filters, rows, columns	In a measure, calculated column, or visual calculation definition	Value in a visual and visual level filter	Value in a visual and visual level filter

reference

Row Context

- Use current row (all row)
- Custom column (M language)
- Calculated columns

```
1   Sales[Price] * Sales[Tax]
```

Filter Context

- Filter before calculation is carried out
- Calculated measures
  - Aggregates on the calculated columns

```
SUM(Sales[Profit])

SUMX(<table>, <expression>)
SUMX(Sales, Sales[Price] * Sales[Tax])
SUMX(FILTER(Sales, Sales[Region]="EMEA"), Sales[Price] * Sales[Tax])

CALCULATE(<expression>, <filter1>, ...other filter conditions)
CALCULATE(SUM(Sales), Sales[Region]="EMEA")
```

VAR and RETURN

- Useful for complax calculation

```
Sales YoY Growth % =
DIVIDE(
    ([Sales] - CALCULATE([Sales], PARALLELPERIOD('Date'[Date], -12, MONTH))),
    CALCULATE([Sales], PARALLELPERIOD('Date'[Date], -12, MONTH))
)
```

```
Sales YoY Growth % =
VAR SalesPriorYear =
    CALCULATE([Sales], PARALLELPERIOD('Date'[Date], -12, MONTH))
RETURN
    DIVIDE(([Sales] - SalesPriorYear), SalesPriorYear)
```

Chapter 8

M language in Power BI

What is M language

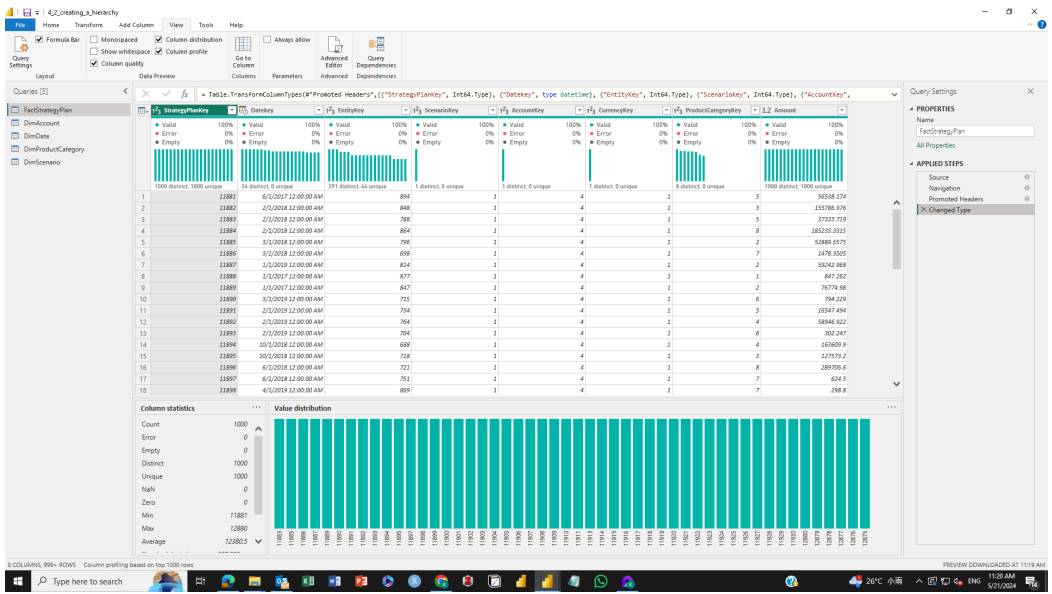
- programme language use in Power Query
- all action in Power Query will be convert to M language
- You can write M language manually in
  - Custom Column
  - Formula bar
  - Advanced Editor
- M language reference
  - https://learn.microsoft.com/en-us/powerquery-m/

Chapter 9

Power Query in Power BI

Data preview

- Quick analyze the data in power query
- Helps diagnose errors and inconsistencies
- Helps you decide what transformation(s) to use



Data Transform

- Pivot column : reshape data for report
- unpivot column: reshape data for data analysis
- Transpose: swap column and row name
- Group by

Chapter 10

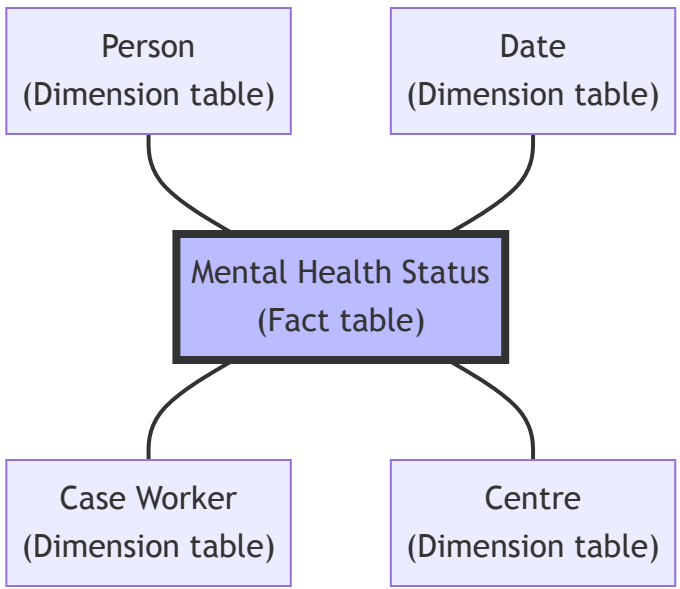
Data Model in Power BI

What is data model

- Tables
- Columns
- Data types
- Relationship between tables
- Keys

Star Schema

- Facts table: metrics
  - Facts (measures)
  - Keys: use to build relationship between tables
- Dimensions table: context
- Benefit
  - Reduce file size
  - Reduce redundant data
  - Easy to manage



Fact table

- Make up of
  - Facts (measures)
  - Keys: use to build relationship between tables
- Fact tables are long and narrow
  - Lots of rows
  - Fewer columns

id	centre_id	depression_scure
1	1	10
2	1	15
3	2	12
4	2	19
5	3	20

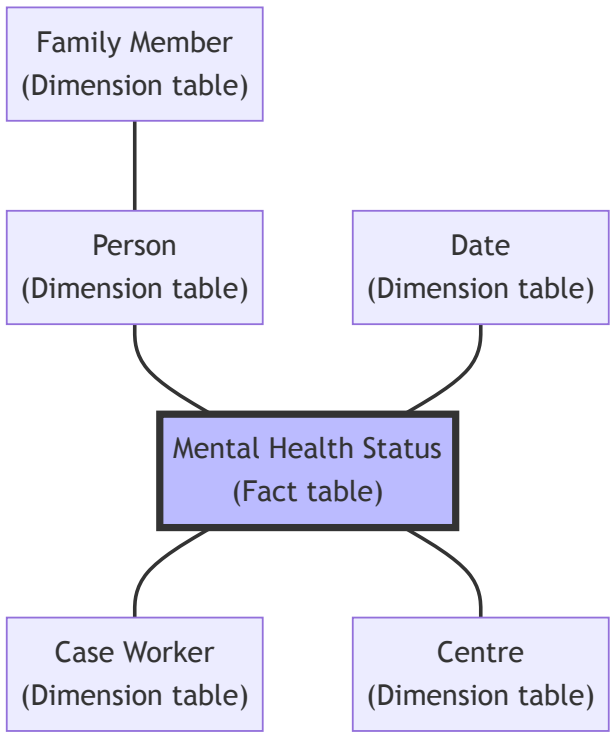
Dimension table

- Provide context
- Shared concepts
- Contain static data
- Dimension tables are short and wide
  - Few of rows
  - Lots of columns

centre_id	name_centre	district_centre
1	ICCMW (Kwai Tsing)	KT
2	ICCMW (Wan Chai)	WC
3	ICCMW (Eastern)	HKE

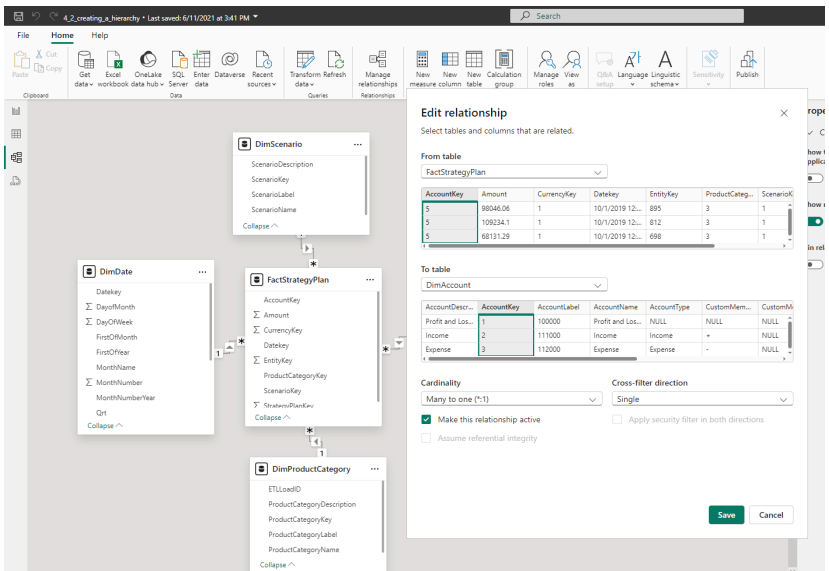
Snowflake Schema

- Allows relationships between dimensions
- Fact table(s) remain the same
- Benefit
  - Reduce file size
  - Reduce redundant data
  - Easy to manage



Build data model in Power BI

- Power BI will define relationship automatically
- You can also define relationship between tables in Model View
  - One-to-many relationship
  - Many-to-many relationship
  - One-to-one relationship
- Each pair of table can have more than one relationship



Build data model in Power BI

- If more than one relationships
  - Set the active relationship in Model View
  - using DAX to activate the relationship when calculating a measure

```
= CALCULATE(SUM(InternetSales[SalesAmount]), USERELATIONSHIP(InternetSales[ShippingDate], DateTime[Date]))
```

Chapter 11

Advanced use case in Power BI

Advanced use case in Power BI

- Button and action
- Mobile View
- Publish
- Drill-through and tooltip
- Bookmark

Button and action

- Create a App-like experience
- Add action to buttons
- Actions
  - Nagative between pages
  - Filter
  - Bookmark

Chapter 12

Integration with Other Microsoft Product

Integration

- Microsoft Power Automate
- Microsoft Forms
- Microsoft Fabric
- Microsoft Power App (Less use in our case)

Microsoft Forms + Power Automate

1. Collect data using Forms
2. Data of Microsoft Forms -> Excel (Power Automate)
3. Add workflow to the Forms
    - Send email
    - Make approval
    - Add new column
    - Conditional control

Microsoft Fabric

1. Excel data -> Lakehouse (Dataflow)
2. Power BI or Notebook
3. Create Scorecard / Alert
4. Trigger another workflow using Power Automate