

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



A word cloud featuring various terms related to data and technology. The words are arranged in a non-uniform, overlapping manner. The colors used are shades of orange, red, purple, and dark blue. The sizes of the words vary, with 'assessment' and 'name' being among the largest. The terms include: assessment, name, calendar, happiness, address, preference, app, distance, instagram, feedback, sleep, facebook, gender, email, attendance, opinion, and purchase.

assessment
name
calendar
happiness
address
preference
app distance
instagram
feedback sleep
facebook
gender
email
attendance
opinion
purchase

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

- Numerical data
- Count, measure, percentage

Examples

- Age
- Temperature

Qualitative data

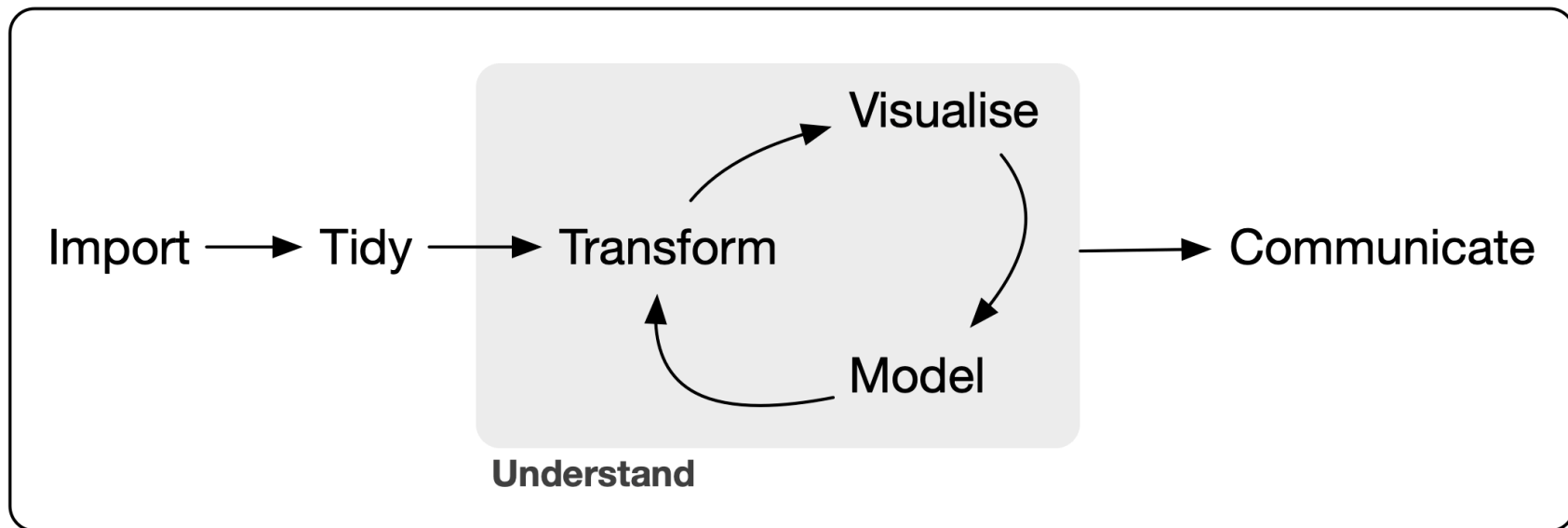
- Categorical data
- Group into categories

Examples

- Gender
- Holiday

Workflow of Data Science

“R for Data Science” (Wickham and Grolemund 2017)



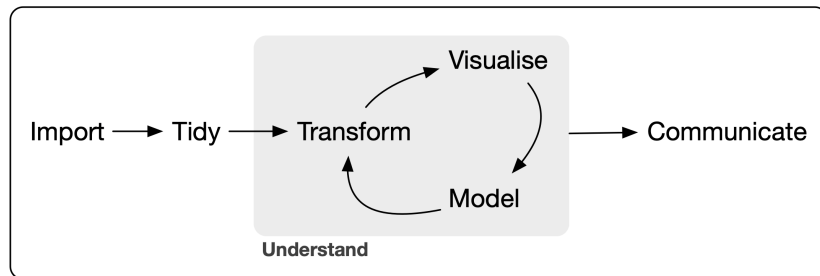
Workflow of Data Science

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



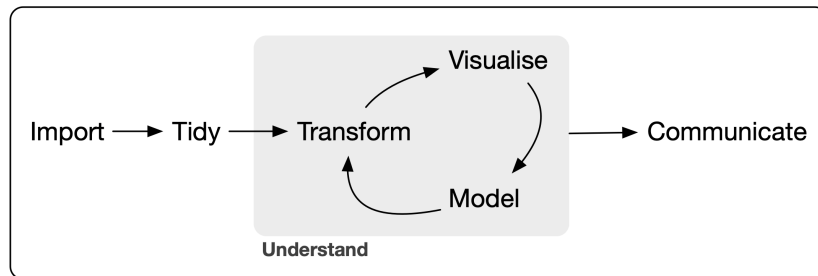
Workflow of Data Science

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



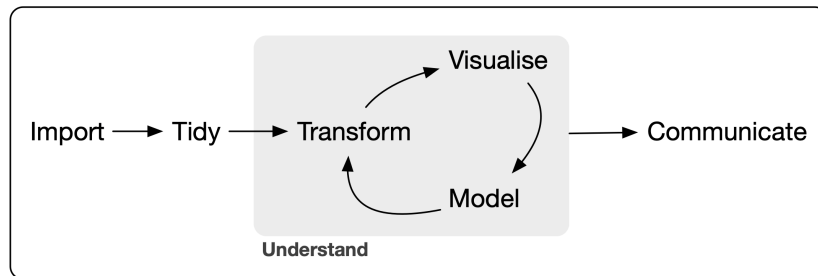
Workflow of Data Science

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



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

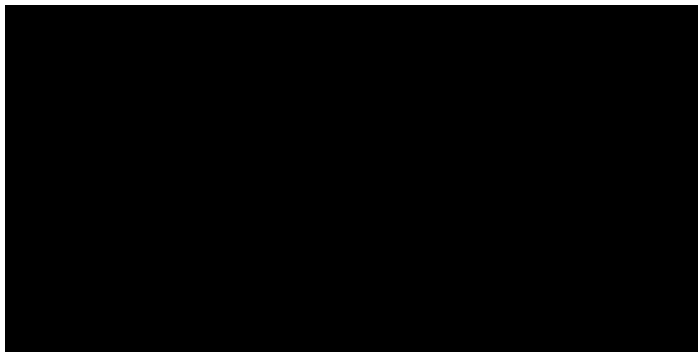
- You can only use the same method to copy and paste

Referencing

Absolute referencing

- **constant** when copied to other cell

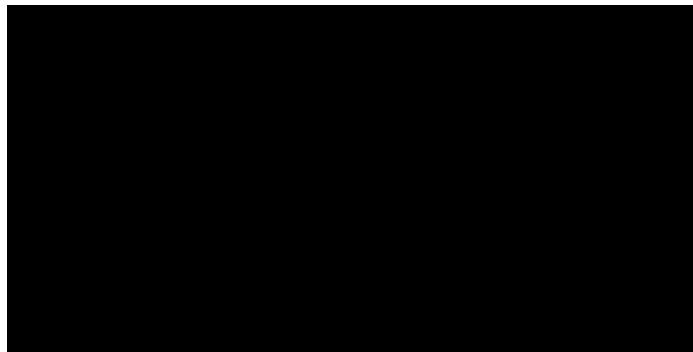
1 =\$A\$2 / =A\$2 / =\$A2



Relative referencing

- **change** when copied to other cell

1 =A2 / =Sheet1!A2

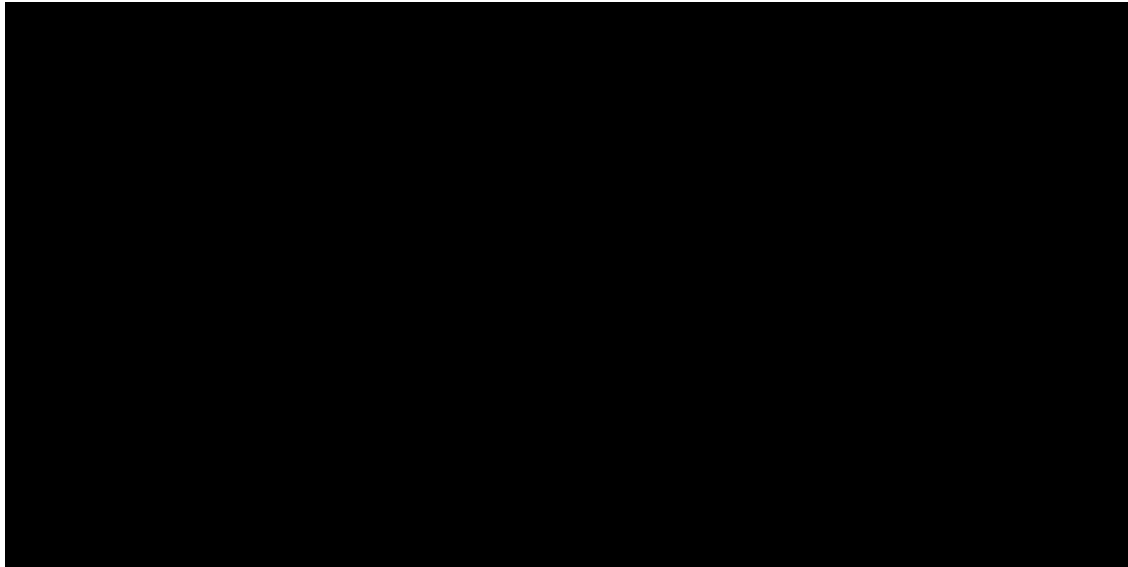


Other topic

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

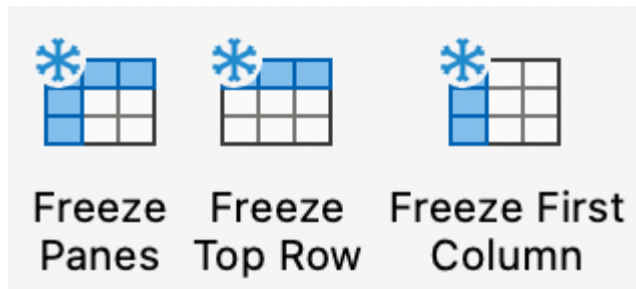
Multiple lines in one cell

- Press Alt + Enter



Freeze cell

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

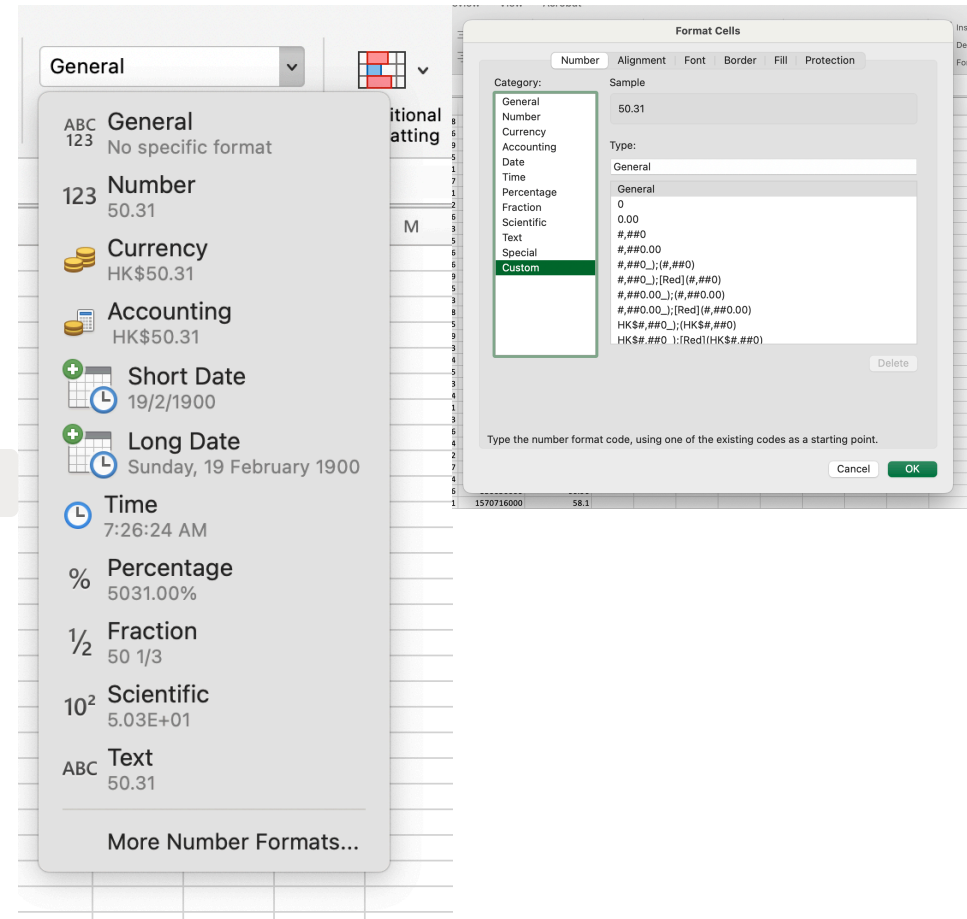


Formatting

- Build-in format
- Custom format
- Using formula

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

reference-date-format



F4 Hotkey

- Make previous action

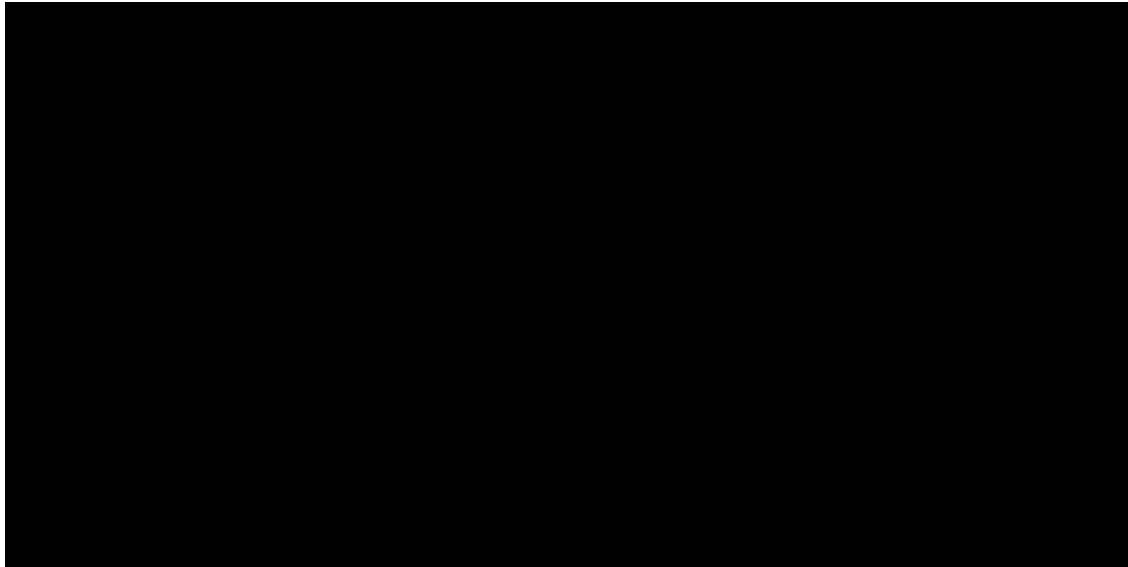


- Change referencing style



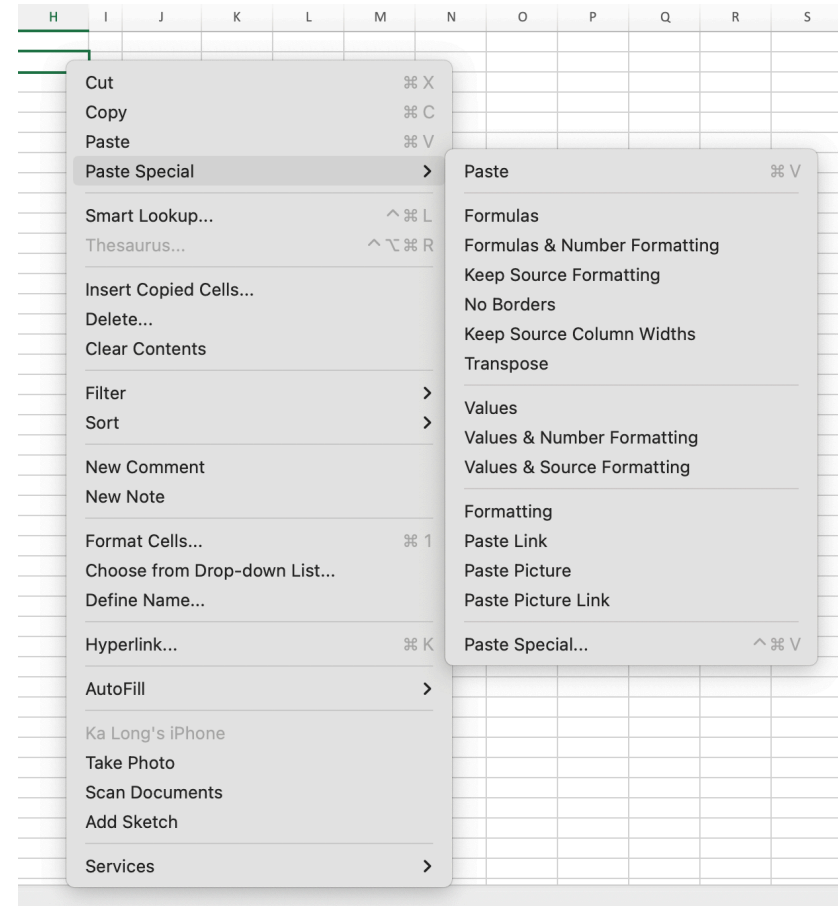
Resize width and height

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

```
1 =[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

- Command button in Ribbon
 - Data validation
 - Conditional formatting
 - Remove duplicates
 - Flash fill
 - Split text to columns
- Operations
- 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

```
1  =SUM(A2:A10)
2  =AVERAGE(A2:A10)
3  =MAX(A2:A10)
4  =MIN(A2:A10)
5  =COUNT(A2:A10)
6  =COUNTBLANK(A2:A10)
```

Text function

```
1  =LEFT()  
2  =RIGHT()  
3  =TRIM()  
4  =CLEAN()  
5  =CONCAT()  
6  =CONCATENATE()  
7  =TEXTJOIN()  
8  =TEXTSPLIT()  
9  =UPPER()  
10 =LOWER()  
11 =PROPER()  
12 =LEN()  
13 =REPLACE()  
14 =SUBSTITUTE()
```

Date & Time function

Date

```
1  =DATE()  
2  =YEAR()  
3  =MONTH()  
4  =DAY()  
5  =DAYS()  
6  =TODAY()  
7  =WEEKDAY()  
8  =WEEKNUM()
```

Time

```
1  =TIME()  
2  =NOW()  
3  =HOUR()  
4  =MINUTE()  
5  =SECOND()
```

Maths function

```
1  =ROUND()  
2  =ABS()  
3  =INT()
```

Logical function

```
1  =AND()  
2  =OR()  
3  =NOT()  
4  =ISNUMBER()  
5  =ISERROR()  
6  =ISERR()  
7  =ISBLANK()  
8  =IF()  
9  =IFS()  
10 =IFERROR()  
11 =SWITCH()
```


Logical + Aggregate function

[Aggregate][Logical]()

```
1 =COUNTIF()  
2 =COUNTIFS()  
3 =SUMIF()  
4 =SUMIFS()  
5 =AVERAGEIF()  
6 =AVERAGEIFS()
```

D[Aggregate]()

```
1 =DMAX()  
2 =DMIN()
```

Lookup function

```
1  =VLOOKUP()  
2  =XLOOKUP()  
3  =FILTER()
```

Array function

Return more than one cell

```
1  =FILTER()  
2  =SORT()  
3  =SORTBY()  
4  =UNIQUE()  
5  =SEQUENCE()
```

Other useful function

```
1  =XMATCH()  
2  =INDEX()  
3  =INDIRECT()  
4  =OFFSET()  
5  =LARGE()  
6  =SMALL()  
7  =ROW()  
8  =ROWS()  
9  =COLUMN()  
10 =COLUMNS()  
11 =CHOOSE()  
12 =SEARCH()  
13 =FIND()
```

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 wildcards

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

big-data-etl

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

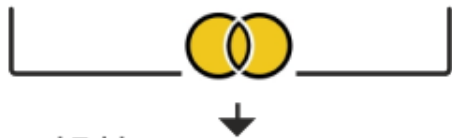
Full join

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

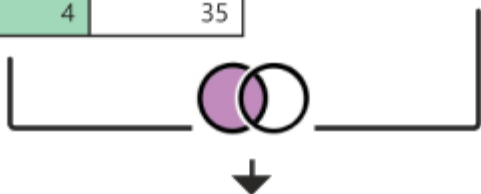
Left join (Right join)

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	<i>null</i>

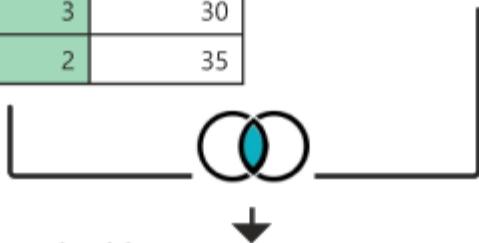
Inner join

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

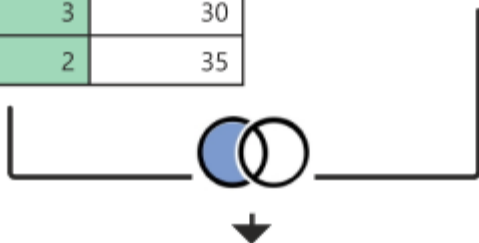
Anti join

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	<i>null</i>
1/2/2020	1	25	<i>null</i>
1/4/2020	2	35	<i>null</i>

Make your decision