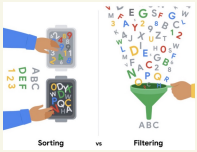
Week 1: Organizing data in Spreadsheet and SQL

Analysis → the process used to make sense of the data collected.
→ Goal: identify trends and relationships within data, so that you can accurately answer to questions you are asking.

→ **"The 4 Phases of Analysis"** (จัดข้อมูลใหม่ / จัดระเบียบข้อมูล / การรู้จักวาง) — make it easier to search, understand e.g. orgnize data into a table, hide/show column

① Organize data → observing and organizing data in a way that's easy to reference.
② Format & Adjust data → adjusting, make it more easy to digest like SORT & FILTER
→ finding Outliers (data points that are very different from similarly collected data & might not be reliable values.)



- **Sort** → arrange data into meaningful order (ASC, DESC), group similar data (→ pivot can sort Monday, Tuesday, ..., Sunday)
- **Filter** → seeing data that meets a specific criteria

③ Get input from others (เปรียบเทียบ / เรียนรู้จากผู้อื่น นำ insight) → checking others' analysis e.g. who experiences on this.
④ Transform data → identify patterns, calculations (e.g. correlation relationships between 2 variables),

Spreadsheet Workshop Commands

Sort
- ✓ Sort sheet (all sheet) * more often use * → all of the data in a spreadsheet is sorted by the ranking of a specific sorted column — data across rows is kept together.
- ✓ Sort Range (only the specified column) → Nothing else on the spreadsheet is rearranged besides the specified in a column.
- ✓ Sort Function → =SORT(A2:D6, sorted_by_column, ASC/DESC) (only specified cell area and create sorted range in a new cell (range) 1, 2, ..., n; TRUE/FALSE)
- ✓ Customized Sort Order (sort by the specified conditions/columns) → Sort data in a spreadsheet using multiple conditions. on Data > Sort range > Advanced range sorting ops.

Get to know Sort & OrderBy

SQL Workshop Commands

- ✓ WHERE clause → to filter only values that meet a specific criteria.
- ✓ SORT BY (ASC by default) → ASC or DESC

Week 2: Convert and Format Data
Combine multiple datasets

Incorrectly formatted data can:
- × Lead to mistakes ✗
- × Take time to fix (L) ...
- × Affect stakeholder's decision-making ☹

Spreadsheet Workshop

- ✓ Typecast numbers (Unit Conversion) → tab above (Format > Number)
  - =CONVERT(E2, "mph", "mls") (หน่วยเดิม, หน่วยใหม่)
  - =CONVERT(E2, "C", "F") (Convert cell E2 from celsius to fahrenheit)
- ✓ LEN, LEFT, RIGHT, FIND → Find position of text string in a cell
- VALUE() convert text string to number.

- ✓ Data Validation (Module 4 Process) In Spreadsheet:
  - Add dropdown lists with predefined options.
  - Create custom checkboxes
  - Protect structured data & formulas (Reject input) will protect from mistyping e.g. is valid date
  - Conditional Formatting highlight cells when matches the condition

SQL Workshop

- CONCAT → CONCAT('google', '.com') → CONCAT(): CONCAT_WS(',', 'www', 'google', 'com'); CONCAT_WS(): 'Google.' + 'com'; CONCAT with +
- CAST(~ AS int64)

(Week2 Continued)

Get support during analysis
- Encounter error when calculating time (start time is greater than end time) change start-end to =IFC(end>start, end-start, 1-start+end)
- Ask team
- Ask/Find questions online
  - Best Practices for searching online
    - Thinking Skills (mental model: your thought process & the way you approach a problem)
    - Data Analytics terms (use the right terms when searching)
    - Basic knowledge tools (can apply new with old knowledge)

Week 3: Aggregate data for analysis (VLOOKUP, JOINS, Subqueries)

- **Aggregation** → collecting or gathering many separate pieces into a whole.
  → organizing pieces of data by Average, Min, Max, Sum

Helps:
- Identify trends
- Make comparisons
- Gain insights

3

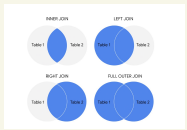
Troubleshooting Questions:
1) How should I prioritize these issues?
2) In a single sentence, what's the issue I'm facing?
3) What sources can help me solve the problem?
4) How can I stop this problem from happening in the future?

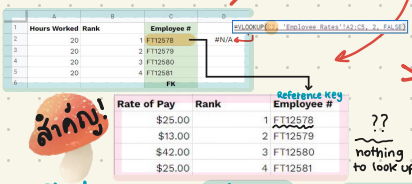
Tools for doing Aggregation:

Spreadsheet
1. VLOOKUP (vertical) (map one product/value to another table)
   - =VLOOKUP(A2, A2:A14, 2, FALSE)
     - search key
     - range
     - index (return column)
     - exact match

* VLOOKUP only looks the data to the right a match is found *
ข้อจำกัด limitation ของ VLOOKUP คือ ต้อง Reference key จะต้องอยู่ก่อน data ที่เรียกหา เท่านั้น (ในที่นี้ต้องอยู่ซ้ายสุด)

* You can protect sheet by clicking on Data > Protected sheets & ranges * to protect the entire sheet / cell from editing



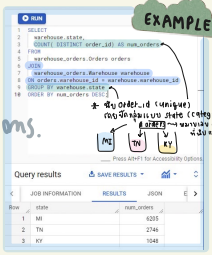


สำคัญ!

Fix by using INDEX() or MATCH()

SQL
- JOINS
  - (INNER) JOIN (by default)
  - LEFT JOIN
  - RIGHT JOIN
  - FULL OUTER JOIN (fetch all records)
- COUNT
  - COUNT
  - COUNT DISTINCT (* not count duplicates *) return distinct values in that range.
- * Inner query executes first *
- Subqueries (or Nested Query / Inner Query / Inner Select)
  - HAVING (WHERE first) to aggregate on statistics functions.
  - CASE → return record that matches like IF-THEN
  - Aliasing → easier & shorter to reference.

EXAMPLE:



→ Nested Query

* array = collection of values like price column

Week 4: Performing Data Calculations (Pivot, SQL calculations, Data-validation, Temp tables)

Spreadsheet:

① Functions & Formulas
- COUNTIF(range, condition) — MAXIFS( )
- SUMIF(range, condition, sum_range)
- SUMPRODUCT(range, [range]) → multiply each record horizontally, then sum it up
  e.g. Quantity × Sales × Margin of each row and plus it together

② Pivot Table (row, column, values, filter)

4

SQL:

- GROUP BY( )
- EXTRACT(YEAR FROM start_time) → to extract a part from a given date. from date/date time

* **Types of Data Validation:** → to make data become complete and accurate, secure, consistent
  1. Data Type → check data type
  2. Data Range → check data falls within an acceptable range
  3. Data Constraints → check data meets certain conditions / criteria.
  4. Data Consistency → check that data makes sense in the context of other related data.
  5. Data Structure → check that data follows / conforms to a set structure
  6. Code Validation → check that the application code systematically performs e.g. more than 1 data type allowed

Temporary table → PROS: Don't have to query in large dataset many times when using often.
a database table that is created and exists temporarily on a database server.

on BigQuery
- Global
- Local

- The WITH clause is a type of temporary table that you can query from multiple times. (Like CREATE VIEW on MySQL) that'll be deleted when you end your SQL session.

```
WITH temp_table_name AS (
  SELECT * FROM
  ~~~~~~~~~~~~~~
)
```

- CREATE TABLE statements
- CREATE TEMP TABLE statements
- not supported on BigQuery: SELECT ~ INTO temp_table_name FROM ~

preferred when finishes using :)
* – Dropping Table: removes info in the table and table variable definitions.
– Deleting Table: removes only info in the table.

Time to "share" your data