Day 3: AI Practice Exercises
Practice 1

KOICA-ESCAP Joint Capacity Building Training Programme on
Digital Technologies for Addressing Climate Change Impacts in the Pacific
20 – 30 October 2026
Brisbane, Australia

Won-Bae Na
Professor of Ocean Engineering
Pukyong National University, Republic of Korea
wna@pknu.ac.kr



Part 1. Theory and Examples

Hands-on Data Analytics and AI Integration (Basic Data Processing and Visualization)

Day 3 Practice 1

30-Minute Theory and Examples + 50-Minute Exercise

Learning Outcomes and Session Plan

Learning outcomes

By completing this practice, participants will learn to apply the full AI-assisted data processing workflow to a new synthetic dataset, covering:

- data cleaning and standardization
- formula generation and aggregation
- chart creation
- Human-in-the-loop verification

Session plan

- Theory and examples: 30 minutes
- Guided exercise: 50 minutes

Operational Scenario

- Your ministry received a corrupted inventory log of rainwater harvesting tanks across multiple island provinces.
- Before generating summary reports for leadership, you must clean the formatting errors and set up aggregation formulas in Microsoft Excel.



Institutional Data Security Policy

- Never copy, paste, or upload real citizen names, exact coordinates of sensitive infrastructure, classified or restricted documents, or unannounced internal government databases into a public AI service.
- This exercise uses only synthetic data. Participants must use approved institutional systems and data-handling procedures for real government work.



Core Concept: Programmatic AI Data Preprocessing

- When managing regional data repositories, manually correcting typos, removing extra spaces, and standardizing capitalization can introduce additional transcription errors.
- Instead of manual editing, officials can use Generative AI as an automated data analyst and formula architect:
  - Automated Cleaning: Providing raw markdown/text tables to the AI to execute bulk formatting fixes (e.g., proper casing, stripping hidden spaces, standardizing blank cells).
  - Formula & Code Generation: Prompting the AI to generate exact Excel formulas (SUMIF, TRIM, PROPER, IFERROR) or Python scripts rather than calculating values manually.

Practice 1 Sequential Workflow

- Data analytics and visualization operate through an automated, programmatic workflow chain:

  `[Messy Raw Dataset] ➔ (Step 1: AI Preprocessing Prompt) ➔ [Cleaned Standardized Table] ➔ (Step 2: Formula & Chart Prompt) ➔ [Excel Formulas & Chart Blueprint] ➔ [Human Verification Audit]`

📐 Step 1: Automated Data Preprocessing

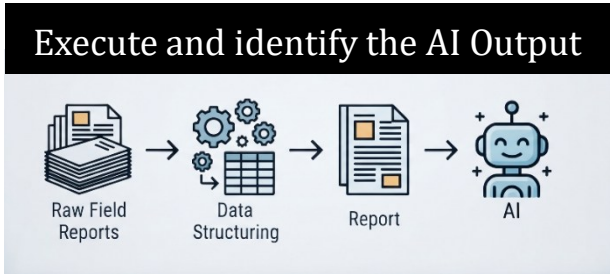
- Paste the prompt 1 into your AI tool alongside the following table:
- Prompt 1:

“Analyze the following markdown table. In the 'Province' column: (1) remove any leading or trailing spaces, (2) replace the misspelling 'providence' with 'Province', and (3) standardize all values to Title Case. For example, ' providence A' and 'province a' must both become 'Province A'. In the 'Tank_Capacity_Liters' column, replace the exact value 'UNKNOWN' with a completely blank cell; do not replace it with 0, 'N/A', or any other text. Preserve all other values exactly as provided. Return only the cleaned table in Markdown format.”

Step 1: Automated Data Preprocessing

- Paste the markdown table after Prompt 1

Site_ID	Province	Tank_Capacity_Liters	Status
TANK-01	providence A	5000	Functional
TANK-02	Province B	UNKNOWN	Operational
TANK-03	province a	3000	Needs Repair



Step 1: Automated Data Preprocessing

- Verify Cleaning: Confirm that ‘ providence A’ and ‘province a’ have both been standardized to ‘Province A’.
- Expected Cleaned Markdown Output:

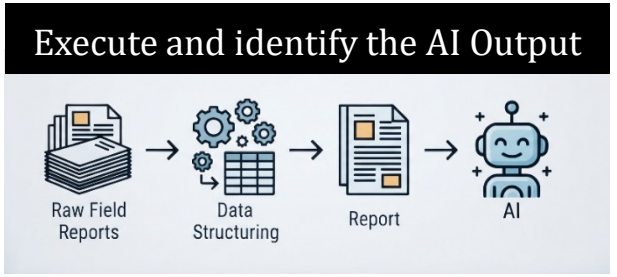
Site_ID	Province	Tank_Capacity_Liters	Status
TANK-01	Province A	5000	Functional
TANK-02	Province B		Operational
TANK-03	Province A	3000	Needs Repair

Key Validation Point: Both ‘ providence A’ and ‘province a’ must be standardized to ‘Province A’.

Step 2: Formula Generation & Chart Architecture

- In the exact same chat thread, type and submit the following visualization instruction:
- Prompt 2:

"Using the parameters of this cleaned dataset, I want to construct a column chart inside Microsoft Excel showing the total combined tank capacity for each individual province. Provide me with the exact Excel formula (such as SUMIF) to aggregate these capacities, along with clear, step-by-step instructions on how to insert and label the bar chart properly."



📐 Step 2: Formula Generation & Chart Architecture

- Expected AI Formula & Chart Output:

  Step 1: Set Up Your Summary Table

  Paste the cleaned table into Excel, beginning in cell A1.

  In a new area of your worksheet, create two-column summary table:

  - Cell F1: Type Province
  - Cell G1: Total Capacity (Liters)
  - Cell F2: Type Province A
  - Cell F3: Type Province B

📐 Step 2: Formula Generation & Chart Architecture

- Continued

Step 2: Enter the Excel SUMIF Formula

To automatically sum the capacities while ignoring the blank/text cells, use the SUMIF formula.

Enter the following formula into cell G2 and press Enter.

```excel
=SUMIF($B$2:$B$4, F2, $C$2:$C$4)
```

Step 2: Formula Generation & Chart Architecture

- Continued

How to apply it to the rest of the table:

1. Click on cell G2.
2. Hover your mouse over the bottom-right corner of the cell until the cursor turns into a small black plus sign (+).
3. Double-click or click-and-drag downward to fill the formula into cell G3.

Step 2: Formula Generation & Chart Architecture

- Continued

Formula Breakdown:

- $B$2:$B$4: The range of cells containing your provinces (locked with $ sings so it doesn’t shift).
- F2: The specific province you are looking for (Province A)..
- $C$2:$C$4: The range of cells containing the tank capacities to be added up. Excel automatically ignores the empty cell in TANK-02.

Step 2: Formula Generation & Chart Architecture

- Continued

Step 3: Insert and Label the Column Chart

1. Select the Summary Table: Highlight your new summary table, including the headers (cells F1 to G3).
2. Insert the Chart:
    - Click on the Insert tab in the top Excel ribbon.
    - In the Charts group, click the Insert Column or Bar Chart icon (the icon with three vertical bars).
    - Select Clustered Column (the first option under 2-D Column).

Step 2: Formula Generation & Chart Architecture

- Continued

3. Add Axis Titles & Labels:

    - Click anywhere inside your newly created chart to select it.
    - Click the green "+" (Chart Elements) button at the top-right corner of the chart design box.
    - Check the box for Axis Titles.
    - Double-click the temporary text on the horizontal X-axis and rename it to Province.
    - Double-click the temporary text on the vertical Y-axis and rename it to Total Capacity (Liters).

Step 2: Formula Generation & Chart Architecture

- Continued

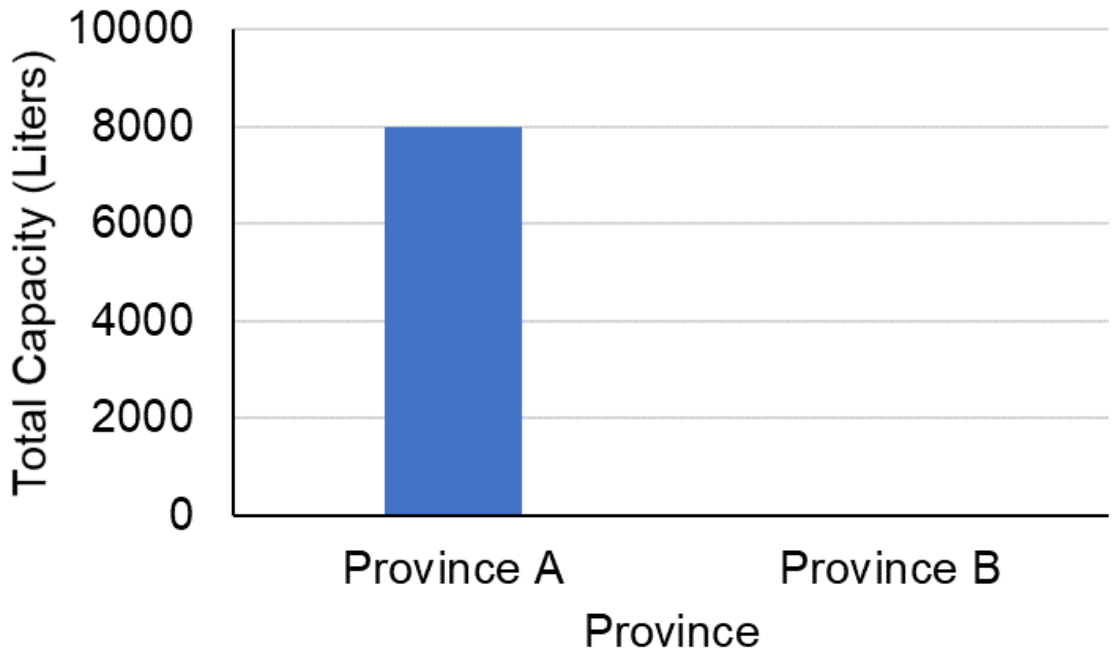
4. Clean up the Title: Double-click the default Chart Title at the top and type a descriptive name, such as Total Tank Capacity by Province.

If you want to customize your visualization further, let me know if you would like to:

- Add Data Labels directly on top of each bar to display exact values.
- Format the numbers in your chart so they include comma separators (e.g., 8,000 instead of 8000) …

Step 2: Formula Generation & Chart Architecture

- Expected Chart (D2_P1_Excel file)



Step 3: Human-in-the-Loop Verification Audit

- Before using these formulas and summary graphics in official reporting, complete the audit checklist below:

Audit Checklist Item	Verification Question	Status (Pass / Fail)
Casing and Trim Check	Did the AI successfully convert all province entries to Title Case without extra spaces (e.g., `Province A`)?	[ ] Pass [ ] Fail
Null Value Check	Was `'UNKNOWN'` properly replaced with a blank cell instead of zero or a text error?	[ ] Pass [ ] Fail
Formula Logic Check	Does the generated `=SUMIF()` formula correctly reference the Province range and Capacity range?	[ ] Pass [ ] Fail

Step 3: Human-in-the-Loop Verification Audit

- Continued.

Audit Checklist Item	Verification Question	Status (Pass / Fail)
Integrity Check	Are the numeric capacities (5000 and 3000) preserved accurately without random alteration?	[ ] Pass [ ] Fail
Spelling/Standar-dization Check	Did the AI correctly fix the spelling error and standardize all province names (e.g., "`providence A`" and "`province a`" to "`Province A`")?	[ ] Pass [ ] Fail



Part 2. Guided Exercise

Hands-on Data Analytics and AI Integration (Basic Data Processing and Visualization)

Day 3 Practice 1

30-Minute Theory and Examples + 50-Minute Exercise

📐 Task 1: Create a Synthetic Dataset

- Create a table containing:
  1. 3 – 12 data rows
  2. At least three provinces or administrative areas
  3. The following four columns: `Asset_ID`, `Province`, `Capacity_Liters`, `Status`
- Intentionally includes at least four data-quality problems:
  1. One value with incorrect capitalization
  2. One value with leading or trailing spaces
  3. One misspelled province name
  4. One `UNKNOWN` capacity value
  5. Optional: one inconsistent status value

Task 1: Create the Raw Dataset

- Example

Asset_ID	Province	Capacity_Liters	Status
ASSET-01	province a	5000	Functional
ASSET-02	Provnce B	UNKNOWN	Operational
ASSET-03	Province C	2500	needs repair

Save an unchanged copy of the raw dataset before cleaning it.

📐 Task 2: Submit the Cleaning Prompt

- Provide the AI the suggested prompt with the raw table and a list of approved province names.
- **Suggested Prompt**:

```
“Analyze and clean the following synthetic dataset. In the
'Province' column, remove leading and trailing spaces, correct
misspellings by mapping each value to the approved province list,
and standardize all values to Title Case. In the 'Capacity_Liters'
column, replace the exact value 'UNKNOWN' with a completely blank
cell; do not replace it with 0, 'N/A', or any other text. Preserve
the Asset_ID values, row order, status values, and all valid
numeric values exactly as provided. Return (1) the cleaned table in
Markdown format and (2) a short change log describing each
correction.”
```

Task 2: Submit the Cleaning Prompt

- **Approved Province Names**:

```
Province A
Province B
Province C
```

- Participant Verification: Before accepting the result, confirm that:

  The number and order of rows have not changed.

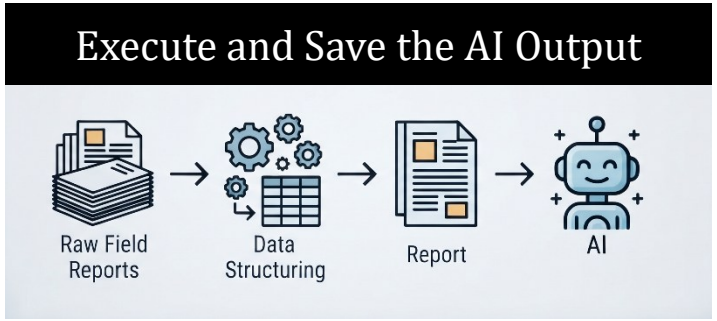
  No `Asset_ID` has been modified.

  Valid capacity values have been preserved.

  All province names match the approved list.

  `UNKNOWN` has been replaced with a blank cell.

- Correct any AI-generated error before proceeding.



📐 Task 3: Build the Summary Table

1. Paste the cleaned table into Excel, beginning in cell A1.
2. Enter Province in F1.
3. Enter Total Capacity (Liters) in G1.
4. List each unique province in F2 and below.
5. Enter the following formula in G2:

   ```
   =SUMIF($B$2:$B$N,F2,$C$2:$C$N)
   ```

   Replace `N` with the final data row number.
6. Fill the formula down for all provinces.
7. Do not manually type or modify the calculated totals.

📐 Task 4: Create the Chart

1. Select the complete summary table, including the headers.
2. Select Insert > Clustered Column Chart.
3. Use the following labels
   - Chart title: `Total Rainwater Capacity by Province`
   - X-axis: `Province`
   - Y-axis: `Total Capacity (Liters)`
4. Confirm that every chart value matches the summary table.

Task 5: Human-in-the-Loop Audit

Participants should mark each item as Pass or Fail.

Audit item	Verification question	Pass / Fail
Data Security Check	Does the dataset contain synthetic data only?	Pass [ ] Fail [ ]
Row Integrity Check	Are the row count and row order unchanged?	Pass [ ] Fail [ ]
Identifier Check	Were all Asset_ID values preserved exactly?	Pass [ ] Fail [ ]
Spelling and Standardization Check	Were spelling errors corrected and all province names standardized?	Pass [ ] Fail [ ]
Null Value Check	Was `UNKNOWN` replaced with a blank cell rather than 0 or text?	Pass [ ] Fail [ ]
Numeric Integrity Check	Were all valid capacity values preserved accurately?	Pass [ ] Fail [ ]
Formula Logic Check	Does the formula reference the correct province and capacity ranges?	Pass [ ] Fail [ ]
Chart Accuracy Check	Do the chart values and labels match the summary table?	Pass [ ] Fail [ ]

If any item fails, correct the problem and repeat the audit.

Submission Requirements

Submit the following items as a single PDF file (including your name and date)

1. Original raw dataset
2. AI-cleaned dataset
3. AI change log
4. Excel summary table (with visible formulas visible)
5. Completed column chart
6. Completed verification checklist
7. A short reflection addressing the following questions:
   - What errors did the AI successfully correct?
   - Did the AI make any unexpected or incorrect changes?
   - What did you verify manually before accepting the result?

Thank you for your attention.

