Capstone Toolkit — Quick Start (Windows)

# Before you start

Download & unzip the project to a simple folder (e.g., D:\capstone-project-cu-gap-analysis\). Avoid syncing/readonly folders (Teams, OneDrive) while running.

Keep the files in the project root:

* ① Install (first-time).bat
* ② Run \_all\_test.bat
* ③ Start Capstone App.bat
* ④ Reset workspace.bat (new: clears previous results before a new dataset)

You only need a normal Windows account. No admin rights required.

# 1) First-time setup (one-time only)

1. Double-click ① Install (first-time).bat.
2. A black window will open and run steps automatically.
3. When you see “OK Environment ready”, close the window (or press any key).

This prepares everything needed to run the tools. You do this only once per computer.

# 2) Run the back end command (one-click with 3 answers)

1. Double-click ② Run \_all\_test.bat.
2. Answer three prompts when asked:

* Value column name (example: Cu\_ppm) — Type the exact column name from your data. Do not add quotes.
* Minimum value (example: 0.05)
* Maximum value (example: 340000)

1. Wait until the script finishes. The window will say “Completed”.

Where are the results? Cleaned CSVs and related outputs are saved under: project\reports\task1\cleaned\ (and related folders shown in the window).

# 3) Open the web app (Streamlit)

1. Double-click ③ Start Capstone App.bat. The first time, it may install Streamlit automatically (internet required).
2. Your default browser should open at: http://localhost:8501/ . If not, open the link manually.
3. To stop the app, return to the black window and press Ctrl + C, then Y.

Optional: If port 8501 is busy, run from Command Prompt: ③ Start Capstone App.bat 8502, then visit http://localhost:8502/

# 4) Reset before analyzing a new ZIP (clears old outputs)

Use the one-click reset:

1. Close the Streamlit window if it is running (Ctrl + C → Y in the black window).
2. Double-click ④ Reset workspace.bat.
3. Choose 1 = Soft reset (only clears reports/…) or 2 = Full reset (also clears the data/ folder).
4. Type YES to confirm (and for Full reset follow any extra on-screen confirmation).
5. When it finishes, copy the new ZIP into data\, then go back to Step 2 (Run \_all\_test).

Manual alternative (if you don’t use the script):

* Delete generated files under reports\task1\ and reports\task2\: \*.csv, \*.parquet, \*.zip and subfolders like cleaned\, diff\, eda\.
* Optionally delete or replace your old ZIP(s) in data\ with the new dataset.
* Do not delete tools\, the .bat files, or start.py.

# Folder map (for reference)

* Project root (this folder): the four BAT files + start.py
* tools\ scripts (internal)
* reports\task1\cleaned\ → cleaned CSVs / outputs
* reports\task2\ → task 2 outputs (if applicable)
* data\ → place your new ZIP here before running Step 2

# Common issues & quick fixes

A. Windows warns “protected PC” (SmartScreen): Click More info → Run anyway. Our scripts are local and safe.

B. The window closes too quickly / I can’t read errors: Our scripts keep the window open. If it still closes, run again and then open run\_log.txt in the project root and tell us what the last 20 lines say.

C. “Value column not found” or similar: Make sure the column name is typed exactly as in your data (e.g., Cu\_ppm). No quotes; watch upper/lower case.

D. Streamlit doesn’t open automatically: Open a browser and go to http://localhost:8501/. If it fails, run the app again and try a different port (see step 3).

E. No results appear in reports\task1\cleaned\: Check the black window (or run\_log.txt) for any error message. Verify the min/max values are reasonable for your dataset.

# Getting help

When contacting support, please attach:

* The file run\_log.txt (found in the project root).
* A screenshot of the black window showing the last lines.
* The three answers you entered (column name, min, max).