CSV Data Quality CLI (Bash/Linux) — Terminal-Bench

One-pager for LinkedIn Media

What it does:

Small Bash CLI that analyzes two CSVs, normalizes CRLF→LF, applies deterministic sorting (LC_ALL=C), detects duplicates, and produces three reports.

How to run:

1) chmod +x solution.sh

2) ./solution.sh   # outputs analysis_report.txt, data_summary.txt, duplicates_found.txt

Outputs:

- analysis_report.txt — counts, basic file info, and tools used (wc, sort, uniq, head, tail).

- data_summary.txt — per-dataset totals and overall summary.

- duplicates_found.txt — per-file duplicate statistics (0 if none).

Engineering notes:

- Deterministic: LC_ALL=C; CRLF→LF normalization; set -euo pipefail.

- Offline test runner; no network installs in run-tests.sh.

- Docker-ready, pinned behavior; clear --help/exit codes (where applicable).

Results:

Local harness checks match assessment style: oracle = 100%, nop = 0%.

Stack:

Bash, Linux coreutils (wc, sort, uniq, head, tail), Docker, Git.

Links:

• Repository: <your GitHub link here>

• README includes examples and usage.