

Stacja 4 (bonus) — porządkowanie niespójnej tabeli

Plik źródłowy: przyklad-korpus/poznan-1956/statystyki/ofiary-i-aresztowania.csv

Tabela jest celowo brudna: różne separatory (; i ,), przecinki dziesiętne, daty w 4 formatach, nagłówki PL+EN, duplikaty, jeden wiersz z **przestawionymi kolumnami**, rozbieżne wartości.

Prompt (wklej, a pod nim wklej zawartość pliku)

Uporządkuj poniższe dane do jednej czystej tabeli. Zasady:

- jeden separator kolumn,
- jeden format daty: RRRR-MM-DD,
- liczby jako liczby (bez „co najmniej”, „ok.” - te przenieś do kolumny „uwaga”),
- usuń duplikaty,
- napraw wiersz, w którym kolumny są przestawione,
- jeśli dwa źródła podają różne wartości tej samej rzeczy, zostaw oba i zaznacz rozbieżność.

Wynik podaj jako tabelę Markdown.

Czego się spodziewać: jednolita tabela; model sam wyłapie duplikat „zabici cywile 50” i wiersz „650;ranni;...” z przestawionymi kolumnami.

Wariant trudniejszy — eksport do czystego CSV

To samo, ale wynik podaj jako poprawny CSV (separator: przecinek, kropka dziesiętna, nagłówki po polsku). Gotowy do wklejenia do Excela / Arkuszy Google.

Po co to w pracy naukowej: dane z różnych źródeł nigdy nie są spójne. To pierwszy krok każdej analizy ilościowej. **Zawsze sprawdź liczby** — model porządkuje format, ale nie jest źródłem prawdy.