

## MAJOR WIN

# Public disclosures into a reproducible, same-rules-for-everyone data pipeline

Public filings → automated reports → reliability flags, validated scan handling, and a methodology — no editorializing

A specialist named **Politics is Good Business** turned public House and Senate financial-disclosure filings into **747** per-member HTML and PDF reports and comparison tables — fully automated, one method for every member, and no commentary layer. This white paper describes the method and the outcomes, not any individual.

## THE PROBLEM

Financial-disclosure filings are public but hard to use: values are reported as ranges, a large share of filings are scanned images, and any comparison can be misread as performance or wrongdoing. The challenge was to make the data readable and comparable for the general public without hiding uncertainty, favouring anyone, or editorializing.

## APPROACH

- 1 Public data → reproducible pipeline** — Fetch index and filings from public sources with paced, identified requests; hash and log every document; regenerate every output from scripts, nothing hand-entered.
- 2 Same rules for every member** — One calculation method and one set of eligibility thresholds; members without enough comparable years are labelled limited or insufficient data, never silently estimated.
- 3 Uncertainty shown, not hidden** — Low / Mid / High bounds from reported ranges and reliability flags (few years, low coverage, unparsed or open-ended values, single-position effects, OCR-derived years). A reliability-filtered recommended list sits beside the full list.
- 4 Scanned filings, with validation** — Two independent readers must agree on a value; conflicts are dropped, not guessed. Checkbox-style forms use a position decoder cross-checked by a second decoder. A scanned filing enters a table only if it clears explicit gates; thresholds are calibrated on electronic filings.

- 5 Methodology and honest limitations** — Methodology and limits are printed in every report: ranges not exact values, residence and some retirement accounts not covered, change in net worth includes salary and gifts, not only investing.
- 6 Disclosure footer, no ads, no trackers** — Every page and PDF carries a free-media-resource disclosure and a not-financial-advice note; no ads, donations, or tracking code; portraits are U.S. government public-domain images with credit.

**Neutral by design:** figures are estimates derived from reported value ranges. They are not financial advice and do not measure investment skill or show wrongdoing. This paper names no individuals and publishes no per-person figures.

## MEASURED OUTCOMES (AGGREGATE)

| Metric                                                      | Value                                                  |
|-------------------------------------------------------------|--------------------------------------------------------|
| Per-member HTML reports                                     | 747                                                    |
| Per-member PDF reports                                      | 747                                                    |
| Percentage table: ranked / limited data / insufficient data | 480 / 112 / 155                                        |
| Net-worth range results (not available)                     | 727 (20)                                               |
| Recommended dollar-change list (reliability-filtered)       | 550                                                    |
| Public-domain portraits embedded                            | 744                                                    |
| Chambers covered                                            | House and Senate (electronic annual reports)           |
| Scanned-filing handling                                     | Validated gates; failures labelled “not used”          |
| Build window (CT)                                           | 2026-09-30 → 2026-10-01                                |
| Status                                                      | Reports and rankings built; public site launching soon |

## WHY IT MATTERS

The trust layer is the product. Identical rules, visible reliability flags, validated handling of hard scanned inputs, and plainly stated limits let readers judge the data themselves — instead of being handed a conclusion. Known gaps (unreadable layouts, paper Senate reports, image-only transaction reports) are listed, not hidden.

## TAKEAWAYS

- Turn public records into public numbers with an automated, reproducible pipeline.
- Apply identical rules to everyone and label what the data cannot support.
- Ship reliability flags, methodology, and limitations with the data, not in a footnote nobody reads.
- Accept machine-read scans only when independent readers agree and validation passes.
- Keep it neutral: no editorializing, no ads, no donations, no trackers.

---

**Privacy note:** Aggregate counts and method description only. Withheld: names of individuals, any per-person figure or rank, filing identifiers, local paths, server details, credentials, and raw transcript text. Agent transcript telemetry was not ingested; none is claimed.

© 2026 SmarTekAI · <https://smartekai.net> · Patterns over voyeurism