Pythia™

Methodology Briefing

Website Digital Health Scoring — How Pythia Works

1. What Is Pythia?

Pythia is a website intelligence scanner that produces a standardised, quantitative assessment of a public-facing website across five dimensions: performance, security, privacy, sustainability, and infrastructure. The output is a single composite score — the **P-Score** (0–100) — and a corresponding letter rating from C to AAA, designed to give analysts, executives, and agencies a consistent, comparable signal across any set of organisations.

The system was built to answer a specific question: can the quality of an organisation's public digital presence serve as a leading indicator of broader operational and governance standards? The hypothesis is that organisations which invest carefully in their digital infrastructure tend to apply similar diligence elsewhere — and conversely, organisations with poor security headers, excessive third-party trackers, or degraded performance may be signalling systemic underinvestment in technical risk management.

2. Data Sources & Methodology

Each scan draws on three independent data sources, combined to produce the final P-Score. The current implementation uses a multi-run averaging approach for maximum stability and accuracy.

2.1 CrUX — Chrome User Experience Report (Field Data)

CrUX is a dataset maintained by Google, built from anonymised, aggregated telemetry collected from real Chrome browser sessions across the web. Where available, it provides the p75 (75th-percentile) value for key loading metrics, meaning 75% of real users experienced a load time at or below the reported figure.

CrUX data is only available for origins with sufficient Chrome traffic to meet Google's privacy threshold. Many institutional and B2B sites do not meet this threshold. The absence of CrUX data does not penalise the score — it simply changes the data source for the performance index. Pythia automatically follows HTTP redirects to ensure it queries the correct final origin.

2.2 Lighthouse — PageSpeed Insights (Lab Data) with 3-Run Averaging

Lighthouse is Google's open-source website auditing engine, run here via the PageSpeed Insights API. It simulates a desktop Chrome browser loading the target URL under controlled conditions and produces scores for performance, accessibility, best practices, and SEO.

3-Run Averaging: To account for the inherent ±2–5 point variance in Lighthouse scores (caused by network variability, server response time fluctuations, and third-party script timing), Pythia runs Lighthouse three times per site and reports the arithmetic mean. Individual run scores are logged for transparency. Runs are sequential with a 2-second cooldown between each.

Lighthouse also surfaces an 'opportunities' list — specific technical issues (unused JavaScript, unoptimised images, etc.) ranked by their estimated impact on load time. These appear in each individual site report and inform the Priority Roadmap section.

2.3 Pythia Proprietary Scan (Headers, HTML & Load Time)

The third data source is Pythia's own HTTP response analysis, which fetches the page and inspects its response headers and HTML content. This scan powers the security, privacy, sustainability, and infrastructure indices and does not rely on any third-party API.

Three-Tier Resilience Strategy

Because institutional and enterprise sites frequently apply bot-blocking controls, the proprietary scan uses a tiered approach before declaring a data failure:

- **Tier 1 — Full GET:** GET request with rotating User-Agents (Chrome, Safari, Googlebot). If any succeeds, full HTML and header data is available — all four proprietary indices are scored.
- **Tier 2 — HEAD Fallback:** If all GET attempts are blocked or time out, a HEAD request is attempted. HEAD requests are rarely blocked as they are used by infrastructure monitoring tools. If HEAD succeeds, response headers are available — Security and Infrastructure can be scored, but Privacy and Sustainability (which require page HTML) are recorded as unavailable.
- **Tier 3 — Complete Failure:** If even HEAD fails, all four proprietary indices are recorded as unavailable.

The table below summarises which indices are scoreable under each scan outcome:

Scan Result	Security	Privacy	Sustainability	Infrastructure
Full GET	✓ Scored	✓ Scored	✓ Scored	✓ Scored
HEAD Only	✓ Scored	N/A	N/A	✓ Scored
Complete Fail	N/A	N/A	N/A	N/A

Load Time Fallback

During a full GET scan, Pythia measures the complete page load time and page weight in MB. If all three Lighthouse runs fail (which is rare), the measured load time serves as a fallback performance score using the formula: **score = max(0, 100 − (load_time × 10))**. A 1-second load time = 90/100; a 2-second load time = 80/100. This ensures every site receives a performance value based on real measured data rather than a synthetic estimate.

3. The Five Indices

The P-Score is a weighted composite of five sub-indices, each scored 0–100:

Index	Weight	Range	What It Measures
Performance	40%	0–100	Lighthouse performance score (3-run average), incorporating Core Web Vitals: LCP, FCP, CLS, TBT, Speed Index. Falls back to measured page load time if Lighthouse is unavailable.
Security	20%	0–100	Presence and configuration of HTTP security headers: HTTPS, HSTS, CSP, X-Frame-Options, X-Content-Type-Options, Referrer-Policy, Permissions-Policy.
Privacy & Tracking	20%	0–100	Detection of third-party tracking technologies in page HTML: analytics (Google Analytics/GTM), session recorders (Hotjar, FullStory, Clarity, etc.), ad trackers, and social media pixels.
Sustainability	15%	0–100	Carbon cost of page delivery: page weight (MB), CDN usage, modern image formats (WebP/AVIF), and lazy loading implementation.
Infrastructure	5%	0–100	CDN deployment (Cloudflare, AWS CloudFront, etc.) and cache header configuration (max-age, public, immutable flags).

4. P-Score Calculation & Rating Scale

4.1 Standard Calculation

When all five indices are available, the P-Score is:

P-Score = (Performance × 0.40) + (Security × 0.20) + (Privacy × 0.20) + (Sustainability × 0.15) + (Infrastructure × 0.05)

Performance is weighted most heavily because it most directly reflects the engineering investment applied to the site. Security and Privacy are equally weighted, reflecting that both represent meaningful compliance and governance signals. Sustainability and Infrastructure carry lower weights as they are more easily influenced by infrastructure provider choice than by active engineering effort.

4.2 Partial Data & Weight Renormalisation

Where one or more indices are unavailable (due to bot-blocking at the proprietary scan stage), Pythia excludes those indices from the calculation and renormalises the remaining weights to sum to 100%, ensuring the P-Score reflects only confirmed data rather than penalising the site with a zero.

For example, if a site blocks all HTTP requests and only Lighthouse data is available, the P-Score is derived from Performance alone (renormalised to 100%). The report clearly flags this limitation: **"Partial data — P-Score based on: performance"**, so readers can calibrate accordingly.

This approach prioritises accuracy and transparency over completeness. A partial-data score is directionally valid and honestly communicated; a score inflated or deflated by unavailable data would not be.

4.3 Rating Scale

Rating	P-Score Range	Interpretation
AAA	95–100	Exceptional
AA	90–94	Excellent
A	85–89	Very Good
BBB	80–84	Good
BB	75–79	Above Average
B	70–74	Fair
CCC	65–69	Below Average
CC	60–64	Poor
C	< 60	Critical

5. Technical Notes

5.1 All Real Measured Data

Pythia uses only real measured data. There are no synthetic scores, estimates, or placeholder values. When Lighthouse succeeds, that data is used. When Lighthouse fails, the actual measured page load time serves as a fallback. Both are real measurements from actual HTTP requests to the target site. Where data is genuinely unavailable, it is recorded as N/A and excluded from scoring rather than substituted with a default.

5.2 Score Variance & Stability

Lighthouse scores can vary ±2–5 points between runs due to network conditions, server response time, and third-party script timing. The 3-run averaging approach substantially reduces this noise. Individual run scores are logged so variance can be inspected. Scores should be treated as directional indicators with a typical confidence band of ±3 points.

5.3 CrUX Availability

CrUX data is absent for many B2B and institutional sites, which is expected and normal given Google's traffic threshold for privacy preservation. Its absence does not penalise the score. Pythia follows HTTP redirects before querying CrUX to ensure it checks the correct final origin.

5.4 Bot-Blocking & Enterprise IP Restrictions

A subset of enterprise and financial institution websites maintain IP blocklists that silently drop connections from cloud provider address ranges (GCP, AWS, Azure). When Pythia is run from such an environment, these sites will time out at the proprietary scan stage regardless of User-Agent rotation. In such cases, the three-tier resilience strategy applies (see Section 2.3), and any successfully scored indices are reported transparently.

5.5 Scope

Each scan covers only the specified URL. It does not assess internal portals, authenticated areas, APIs, or backend infrastructure. The P-Score represents the digital health of the public-facing entry point — the face the organisation presents to the world.

5.6 Plain English Reporting

Individual site PDF reports include plain English explanations for every metric — 'Why It Matters' columns, diagnosis sections explaining performance bottlenecks in business terms, and actionable Priority Roadmaps with effort/impact estimates. The intent is to make technical findings legible to a non-technical executive audience without sacrificing analytical rigour.