

Crawl & Index Audit Report

DOMAIN <https://www.example.com>
CRAWL DATE **April 18, 2025**
TOOL **Screaming Frog 20.1 + GSC**
ANALYST **SEO Team**

3,842

Total URLs

2,610

Indexable

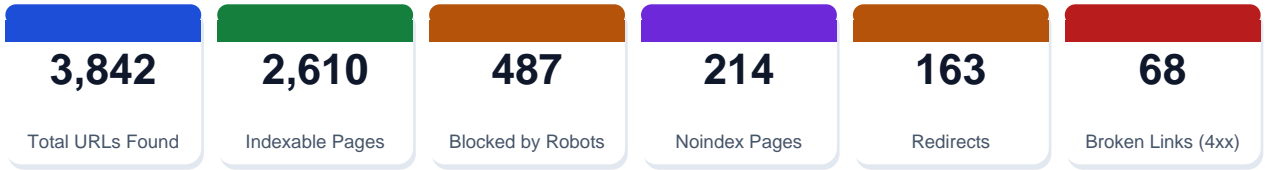
68

Broken (4xx)

163

Redirects

01 Crawl Overview



Indexability Breakdown

Status	Count	% of Total	Notes
Indexable	2,610	67.9%	Eligible for Google index
Noindex — meta tag	214	5.6%	Excluded via meta robots tag
Noindex — HTTP header	38	1.0%	X-Robots-Tag: noindex
Blocked by robots.txt	487	12.7%	Disallow rules prevent crawl
Redirects (3xx)	163	4.2%	Check for unnecessary chains
Client Error (4xx)	68	1.8%	Broken pages — fix or redirect
Server Error (5xx)	12	0.3%	Intermittent server errors
Canonicalised (non-self)	250	6.5%	Pointing to alternate canonical

02 Robots.txt Analysis

Directive	Path / Pattern	User-agent	Assessment
Disallow	/wp-admin/	*	Admin area — correctly blocked
Disallow	/staging/	*	Staging folder blocked — good
Disallow	/search?*	*	Faceted search params blocked
Disallow	/cart/	*	Cart pages — correct
Disallow ■	/blog/	GPTBot	AI crawler blocked — review intent
Sitemap	/sitemap_index.xml	—	Sitemap declared — correct
Crawl-delay	Not set	*	No delay — acceptable for most sites

■ GPTBot is blocked on [/blog/](#). If AI-generated traffic or training exposure is desired, remove this rule.

Sitemap File	URLs	Indexed	Issues	Status
/sitemap_index.xml	—	—	None	VALID
/sitemap-pages.xml	842	831	11 URLs not yet indexed	WARN
/sitemap-posts.xml	1,204	1,198	6 noindex pages inside sitemap	WARN
/sitemap-products.xml	788	782	None	VALID
/sitemap-images.xml	3,210	N/A	None	VALID
/sitemap-categories.xml	64	64	None	VALID

■ 6 noindex pages found inside sitemaps. Remove noindex URLs from XML sitemaps — conflicting signals confuse Googlebot.

Issue	Count	Severity	Example
Self-referencing canonical	2,360	PASS	/products/blue-widget/
Canonical pointing to redirect	28	ERROR	/old-page/ redirects to /new-page/
Missing canonical tag	92	WARN	/tag/blue/ (taxonomy pages)
Canonical targets a noindex page	7	ERROR	/promo-expired/
Cross-domain canonical	14	WARN	m.example.com -> www.example.com
Multiple canonical tags on page	3	ERROR	/checkout/step-2/

05 Redirect Audit

Redirect Type	Count	Status	Recommended Action
301 Permanent Redirect	128	PASS	Preferred — no action needed
302 Temporary Redirect	21	WARN	Change to 301 if permanent
Redirect Chain (3+ hops)	17	REDIRECT CHAIN	Collapse each to a single hop
Redirect Loop	2	ERROR	Immediate fix required
Redirect to a 4xx page	5	ERROR	Destination page is broken
HTTP to HTTPS (site-wide)	All	PASS	Correctly implemented
WWW to non-WWW	All	PASS	Canonical domain enforced

06 Crawl Depth Distribution

Depth	URL Count	% of Crawl	Assessment
0 — Homepage	1	0.03%	—
1	48	1.2%	Good — key sections reachable in one click
2	312	8.1%	Good
3	1,104	28.7%	Acceptable
4	1,568	40.8%	Review internal linking structure
5	621	16.2%	Improve navigation — borderline deep
6+	188	4.9%	High risk — may receive reduced crawl budget

■ Pages beyond click depth 4 risk reduced crawl budget. Flatten site architecture by improving internal linking on key hub pages.

07 Google Search Console — Index Coverage

GSC Status	Count	MoM Change	Notes
Indexed	2,487	+124 (+5.2%)	Core pages visible in Google

Crawled — not indexed	312	+18 (+6.1%)	Google crawled but chose to exclude
Discovered — not indexed	198	-24 (-10.8%)	Queued; not yet crawled
Duplicate — Google chose alt canonical	241	+6 (+2.5%)	Review canonical signals
Excluded by noindex	252	0 (stable)	Intentional exclusions — OK
Page with redirect	163	-12 (-6.9%)	Tracked by GSC
Not found (404)	68	+3 (+4.6%)	Fix or redirect broken pages

08 Duplicate & Thin Content Issues

Issue Type	Count	Severity	Recommended Action
Exact duplicate pages	34	ERROR	Consolidate via canonical or 301 redirect
Near-duplicate content	87	WARN	Differentiate or merge similar pages
Thin content (under 300 words)	156	WARN	Expand content or apply noindex
Pagination duplicates	48	WARN	Use canonical to page 1
URL parameter duplicates	62	ERROR	Configure GSC parameter handling
Print-friendly version dupes	12	WARN	Add noindex to print view pages

09 Crawl Budget Warnings

Issue	URLs Affected	Priority	Fix
Faceted navigation URLs indexed	284	ERROR	Block via robots.txt or noindex
Session / UTM params creating duplicates	118	ERROR	Canonical + parameter stripping
Internal search result pages indexed	73	WARN	Add noindex to search results
Orphan pages — no internal links	41	WARN	Add links or remove pages
Low-value tag and archive pages	96	WARN	Noindex or consolidate
Infinite scroll / AJAX content	N/A	PASS	Pagination implemented correctly

10 Prioritised Action Plan

P1 Critical — Fix Immediately

- Fix 2 redirect loops — these can prevent entire URL chains from being crawled.
- Remove noindex pages from all XML sitemaps (6 instances identified).
- Fix canonical tags pointing to redirected or noindex URLs (35 pages).
- Resolve redirect-to-4xx destinations (5 broken redirect targets).

P2 High — Fix Within 2 Weeks

- Eliminate 62 URL-parameter duplicates via canonical tags or robots.txt.
- Consolidate 34 exact-duplicate pages with 301 redirects.
- Collapse 17 redirect chains to a single hop.
- Fix or redirect all 68 broken 4xx pages.

P3 Medium — Fix Within 1 Month

- Add missing canonical tags to 92 untagged pages.
- Improve or noindex 156 thin-content pages (under 300 words).
- Reduce click depth for 188 URLs sitting at depth 6+.
- Block 284 faceted navigation URLs from being indexed.

P4 Low — Ongoing Improvement

- Change 21 temporary 302 redirects to permanent 301s.
- Noindex 96 low-value tag and archive pages.
- Review GPTBot disallow rule on /blog/ if AI exposure is desired.
- Add internal links to 41 orphan pages or remove them.

Report Summary

OVERALL SITE HEALTH SCORE	CRITICAL ISSUES	WARNINGS	PASSED CHECKS
62 / 100	9	18	14
Needs Improvement	Immediate action required	Fix within 30 days	No action needed