

Google Search Console Indexing Analysis Report

Site: **avinashsangle.com**

Date: February 8, 2026

Framework: Next.js (Static Export) on GitHub Pages

EXECUTIVE SUMMARY

Google Search Console reports **23 pages not indexed** and only **12 pages indexed** with just **9 total clicks**. The biggest category is **'Discovered - currently not indexed' (12 pages)** where Google found your pages via the sitemap but decided not to crawl or index them. The root cause is a **critical canonical URL misconfiguration** in layout.tsx that tells Google every page is a duplicate of the homepage. Additional issues include 5 redirect pages, 3 broken 404s from your old site structure, and 1 canonical conflict.

Category	Status	Impact
Canonical URLs	CRITICAL	12 of 14 pages have wrong canonical pointing to homepage
Google Indexed Pages	CRITICAL	Only ~2 of 17 pages indexed (88% missing)
Sitemap.xml	OK	Valid, accessible, 14 URLs listed
Robots.txt	OK	Properly allows all crawlers
Meta Titles	OK	All pages have unique titles
Meta Descriptions	OK	All pages have unique descriptions
Robots Meta Tag	OK	All pages set to index, follow
JS Bundle Size	MEDIUM	~3.2 MB total JS - may slow crawl budget
Showcase Page	HIGH	Exists but missing from sitemap

1. CRITICAL: Canonical URL Misconfiguration

What's happening: Almost every page on your site has a `<link rel='canonical'>` tag pointing to <https://avinashsangle.com/> (the homepage), instead of pointing to its own URL. This tells Google that all these pages are duplicates of the homepage, so Google ignores them.

Page	Current Canonical	Correct Canonical	Status
/	https://avinashsangle.com/	https://avinashsangle.com/	OK
/blog/	https://avinashsangle.com/	.../blog/	WRONG
/blog/clawdbot-guide/	https://avinashsangle.com/	.../blog/clawdbot-guide/	OK
/blog/method-crm-mcp/	https://avinashsangle.com/	.../blog/method-crm-mcp/	WRONG
/projects/	https://avinashsangle.com/	.../projects/	WRONG
/projects/jenkins-mcp/	https://avinashsangle.com/	.../projects/jenkins-mcp/	WRONG
/projects/reddit-agent/	https://avinashsangle.com/	.../projects/reddit-agent/	WRONG
/projects/calculator-server/	https://avinashsangle.com/	.../projects/calculator-server/	WRONG
/projects/aws-ec2-agent/	https://avinashsangle.com/	.../projects/aws-ec2-agent/	WRONG
/projects/twitter-oauth/	https://avinashsangle.com/	.../projects/twitter-oauth/	WRONG
/projects/method-crm-mcp/	https://avinashsangle.com/	.../projects/method-crm-mcp/	WRONG
/projects/social-media-.../	https://avinashsangle.com/	.../projects/social-media-.../	WRONG
/projects/wp-mcp/	https://avinashsangle.com/	.../projects/wp-mcp/	WRONG
/showcase/	https://avinashsangle.com/	.../showcase/	WRONG

Result: 12 out of 14 pages have incorrect canonicals. Only the homepage and /blog/clawdbot-guide/ are correct.

Root Cause

In `src/app/layout.tsx` (line 70-72), the root layout sets a global canonical to '/' for ALL pages:

```
alternates: { canonical: "/" }
```

Since Next.js resolves this relative to metadataBase (<https://avinashsangle.com>), every page gets canonical = <https://avinashsangle.com/>. Only the clawdbot-guide page overrides this with its own canonical. All other pages inherit the wrong value.

The Fix

Option A (Recommended): Remove the global canonical from layout.tsx entirely:

```
// In src/app/layout.tsx - DELETE these lines: alternates: { canonical: "/", },
```

Next.js with metadataBase will auto-generate the correct canonical for each route.

Option B: Add explicit canonical overrides in each page's metadata export:

```
// Example: src/app/blog/page.tsx export const metadata: Metadata = { ... alternates: {  
canonical: "/blog/", }, };
```

2. Google Search Console: Full Indexing Breakdown

Data pulled directly from your GSC account (last updated Feb 3, 2026). GSC shows **23 not-indexed pages** across 6 different reasons and **12 indexed pages**:

Reason	Source	Pages	Severity
Discovered - currently not indexed	Google systems	12	CRITICAL
Page with redirect	Website	5	HIGH
Not found (404)	Website	3	HIGH
Alternative page with proper canonical tag	Website	1	CRITICAL
Redirect error	Website	1	MEDIUM
Crawled - currently not indexed	Google systems	1	MEDIUM

2a. 'Discovered - currently not indexed' (12 pages)

Google found these URLs in your sitemap but has **not even crawled them yet** (Last Crawled: N/A). This strongly correlates with the canonical issue - Google may be deprioritizing these pages because their canonical tag points to the homepage.

URL	Last Crawled
/blog/	N/A
/blog/method-crm-mcp/	N/A
/projects/	N/A
/projects/aws-ec2-agent/	N/A
/projects/calculator-server/	N/A
/projects/jenkins-chatbot/	N/A
/projects/jenkins-mcp/	N/A
/projects/method-crm-mcp/	N/A
/projects/reddit-agent/	N/A
/projects/social-media-auto-poster/	N/A
/projects/twitter-oauth/	N/A
/projects/wp-mcp/	N/A

2b. 'Page with redirect' (5 pages)

These are old URLs or variant domains that redirect. This is **expected behavior** for the domain variants (http/www redirecting to https). The clawdbot-guide without trailing slash redirects to the trailing slash version.

The old /make-twitter-oauth/ URL is from a previous site structure.

URL	Crawled
http://avinashsangle.com/	3 Feb 2026
https://www.avinashsangle.com/	3 Feb 2026
.../blog/clawdbot-guide (no trailing slash)	31 Jan 2026
http://avinashsangle.com/make-twitter-oauth/	28 Jan 2026
http://www.avinashsangle.com/	15 Jan 2026

2c. 'Not found - 404' (3 pages)

These are **old URLs from your previous site structure** (plain .html files) that no longer exist after the Next.js migration. Google still remembers them and keeps checking.

URL	Crawled
https://avinashsangle.com/blog.html	30 Jan 2026
https://avinashsangle.com/project-aws-ec2-agent.html	30 Jan 2026
https://blog.avinashsangle.com/cdn-cgi/l/email-protection	19 Jan 2026

Fix: Add 301 redirects from old URLs to new ones: /blog.html -> /blog/, /project-aws-ec2-agent.html -> /projects/aws-ec2-agent/. The Cloudflare email protection URL can be ignored.

3. HIGH: Showcase Page Missing from Sitemap

The **/showcase/** page exists as a fully rendered HTML page and is crawlable (robots: index, follow), but it is **not included in sitemap.xml**. Additionally, its title has a duplication bug: "UI Component Showcase | Avinash Sangle | Avinash Sangle" (the name appears twice due to the title template).

Fix: Either add /showcase/ to the sitemap generation, or if it's an internal/dev page, set `robots: { index: false }` in its metadata to prevent crawling.

4. MEDIUM: Large JavaScript Bundles

The site ships approximately **3.2 MB of JavaScript** across multiple chunks. The three largest chunks are each ~549 KB (likely React DOM, Recharts, and Radix UI). While the site is statically exported (so core content is in HTML), large JS bundles can:

- Consume Googlebot's crawl budget and rendering resources
- Slow down page load speed (a ranking factor)

- Delay First Contentful Paint and Largest Contentful Paint metrics

Fix: Consider code-splitting more aggressively, lazy-loading Recharts and Radix UI components, and auditing unused dependencies. Run `npx @next/bundle-analyzer` to identify the largest contributors.

5. What's Working Well

robots.txt: Correctly configured with 'Allow: /' and sitemap reference.

sitemap.xml: Valid XML, accessible at /sitemap.xml, lists 14 URLs with lastmod dates and priorities.

Meta Titles: Every page has a unique, descriptive title using the Next.js title template.

Meta Descriptions: All pages have unique, keyword-rich descriptions (150-160 characters).

Robots Meta Tag: All pages correctly set to 'index, follow'.

Open Graph Tags: Most pages include OG title, description, URL, and images.

Twitter Cards: Homepage and blog posts have proper Twitter card markup.

Structured Data: The clawdbot-guide page is exemplary with TechArticle, BreadcrumbList, HowTo, FAQPage, and SoftwareApplication schemas.

Static HTML: Since the site uses 'output: export', all content is pre-rendered in HTML. Googlebot can read it without JavaScript rendering.

Internal Linking: Navigation and page cross-links use proper <a href> tags.

Trailing Slash Consistency: next.config.ts enforces trailing slashes, preventing duplicate URL issues.

6. Action Plan (Priority Order)

#	Priority	Action	Details	Time
1	CRITICAL	Fix canonical URLs in layout.tsx	Remove 'alternates: { canonical: "/" }' from src/app/layout.tsx. This single change will fix 12 pages. Rebuild and redeploy the site.	~15 min
2	CRITICAL	Resubmit sitemap in Search Console	After deploying the fix, go to Google Search Console > Sitemaps > resubmit sitemap.xml. Then use 'URL Inspection' to request indexing for each page.	~30 min
3	HIGH	Add 301 redirects for old URLs	Redirect /blog.html to /blog/ and /project-aws-ec2-agent.html to /projects/aws-ec2-agent/. Use Next.js redirects in next.config.ts or a _redirects file for GitHub Pages.	~20 min
4	HIGH	Fix showcase page	Either add /showcase/ to sitemap or set its robots to noindex. Fix the duplicate title.	~10 min
5	MEDIUM	Add structured data to all project pages	The clawdbot-guide page is a great template. Add similar JSON-LD schemas (SoftwareApplication, BreadcrumbList) to all project pages.	~2 hrs
6	MEDIUM	Optimize JS bundle size	Use bundle analyzer to identify and lazy-load heavy libraries. Target less than 1.5 MB total JS.	~3 hrs

7	LOW	Monitor indexing progress	Check Search Console weekly for 4-6 weeks after fixes. Google typically re-crawls and indexes within 1-4 weeks.	Ongoing
---	-----	---------------------------	---	---------

EXPECTED TIMELINE AFTER FIXES

Week 1: Deploy canonical fix + resubmit sitemap. Google begins re-crawling.

Week 2-3: Google processes updated pages. Expect 5-10 pages indexed.

Week 4-6: Full indexing of all pages. Search Console coverage report shows green.

Month 2+: Pages start ranking for target keywords. Organic traffic increases.