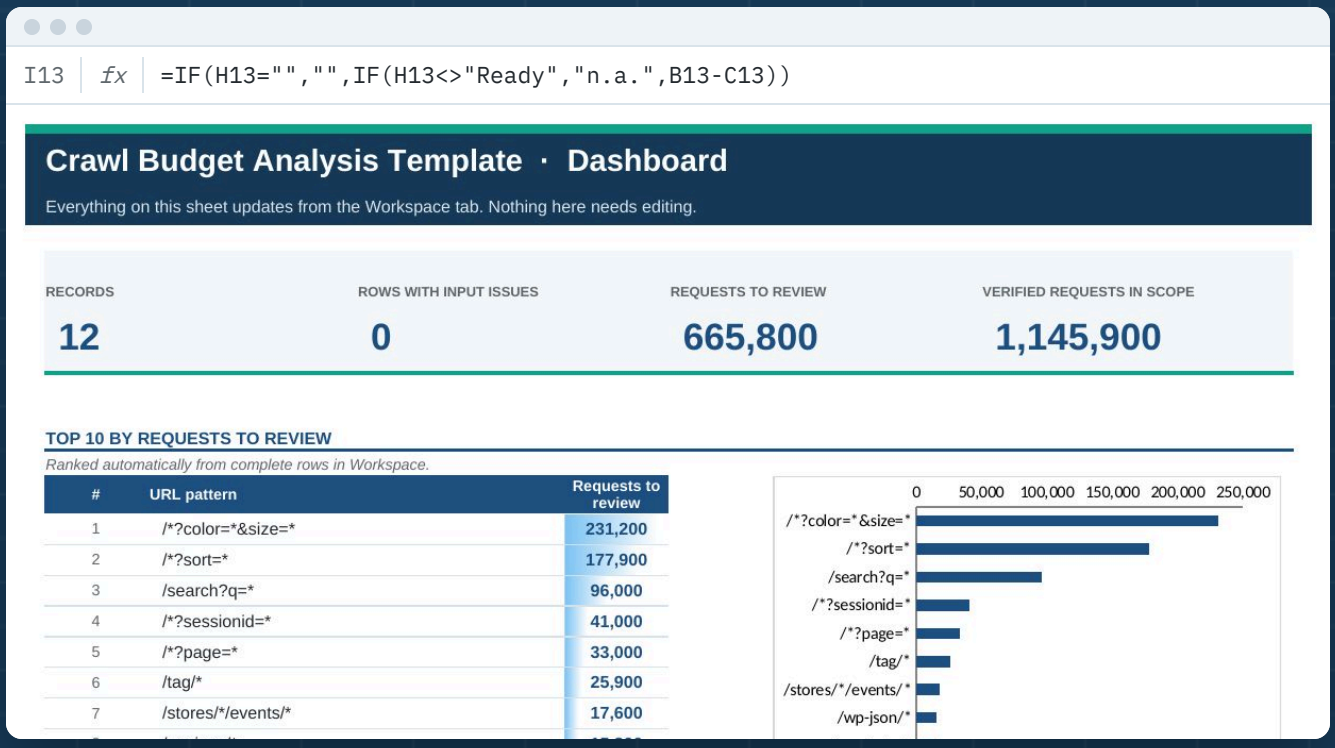
OptimizationTheory.com

Technical SEO Toolkit, resource 14

Crawl Budget Analysis Template

Group verified Googlebot requests by URL pattern, measure crawl waste and pick the right fix for each pattern.

Built for the DataForSEO API: On-Page API



Field guide, v3.0, November 2026

Jourdan Rombough, optimizationtheory.com

What this template decides

Which URL patterns soak up crawl requests without returning value, how much crawl activity each wastes, and what to do about it.

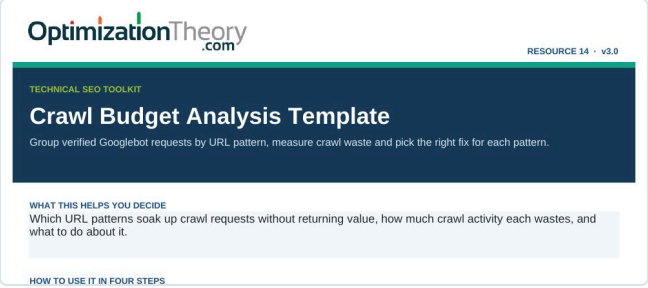
Crawl budget is a real constraint on large sites: ecommerce catalogs with faceted navigation, marketplaces, publishers with deep archives. When Googlebot spends hundreds of thousands of requests on sort orders, filter combinations and session IDs, new and updated pages wait longer to be crawled.

This crawl budget analysis template works from verified Googlebot log data grouped by URL pattern. For each pattern you enter total verified requests and how many hit valuable URLs, plus unique URLs and internal links. It calculates requests to review, the share of the pattern that's waste, and requests per URL, so you can match each pattern to the right fix.

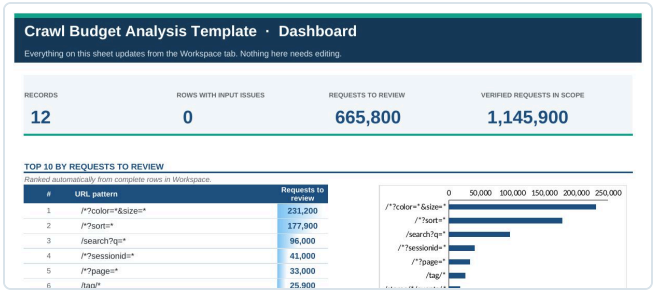
Who it's for

- Technical SEOs on large ecommerce or publisher sites
- Developers deciding how to handle faceted navigation and parameters
- Teams investigating slow indexing of new pages
- Anyone with server logs and a suspicion something's wrong

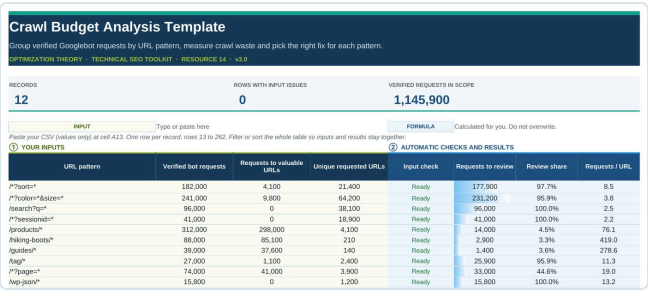
What's in the workbook



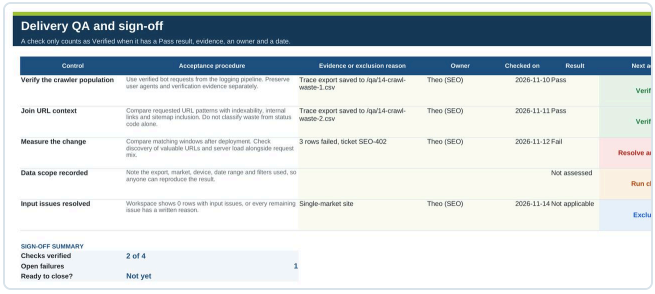
Start Here. Purpose, four-step setup and colour key.



Dashboard. KPIs, charts, Top 10 and an automatic action list.



Workspace. The working table, room for 250 rows.



Delivery QA. Evidence, owner and date for each validation step.

How the calculation works

Requests to review = verified requests – valuable requests. Review share = requests to review ÷ verified requests. Requests per URL = verified requests ÷ unique URLs.

The key formula, as written in the workbook

```
=IF(H13="","",IF(H13<>"Ready","n.a.",B13-C13))
```

Result column	Plain-English formula	What it tells you
Requests to review	Verified bot requests-Requests to valuable URLs	Requests outside the approved valuable subset. Not all are waste.
Review share	IF(Verified bot requests=0,"n.a.",Requests to review ÷ Verified bot requests)	Share of verified requests outside the valuable subset.
Requests / URL	Verified bot requests ÷ Unique requested URLs	Request frequency for this observed window.

How to use it, step by step

1

Verify the bot

Filter logs to verified Googlebot using reverse DNS. User agents alone are easily spoofed.

2

Group by pattern

Bucket URLs into patterns: parameters, folders, templates. Count requests and unique URLs per pattern for a fixed window.

3

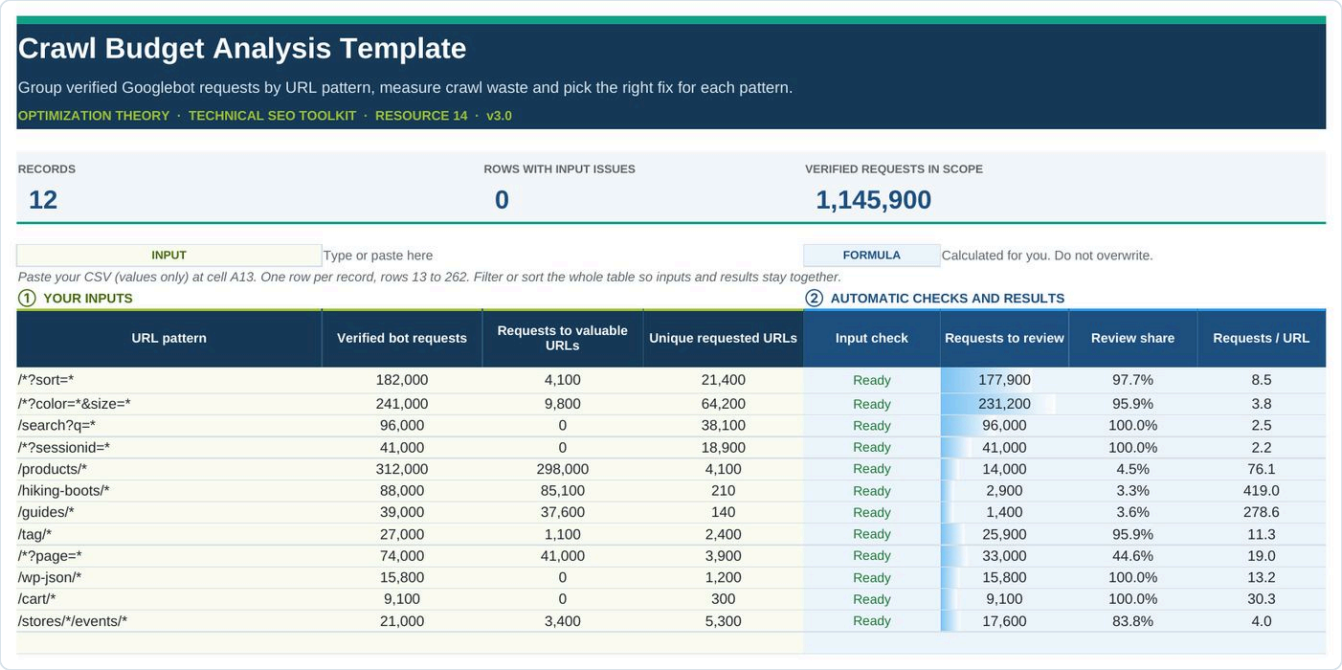
Mark valuable requests

Valuable means the request hit an indexable, canonical URL you want crawled. Join with your crawl data.

4

Choose an action per pattern

High review share with many internal links usually means fixing links; high requests per URL means a trap.



Crawl Budget Analysis Template

Group verified Googlebot requests by URL pattern, measure crawl waste and pick the right fix for each pattern.

OPTIMIZATION THEORY · TECHNICAL SEO TOOLKIT · RESOURCE 14 · v3.0

RECORDS	ROWS WITH INPUT ISSUES	VERIFIED REQUESTS IN SCOPE
12	0	1,145,900

INPUT Type or paste here — FORMULA Calculated for you. Do not overwrite.

Paste your CSV (values only) at cell A13. One row per record, rows 13 to 262. Filter or sort the whole table so inputs and results stay together.

① YOUR INPUTS — ② AUTOMATIC CHECKS AND RESULTS

URL pattern	Verified bot requests	Requests to valuable URLs	Unique requested URLs	Input check	Requests to review	Review share	Requests / URL
/*?sort=*	182,000	4,100	21,400	Ready	177,900	97.7%	8.5
/*?color=*&size=*	241,000	9,800	64,200	Ready	231,200	95.9%	3.8
/search?q=*	96,000	0	38,100	Ready	96,000	100.0%	2.5
/*?sessionid=*	41,000	0	18,900	Ready	41,000	100.0%	2.2
/products/*	312,000	298,000	4,100	Ready	14,000	4.5%	76.1
/hiking-boots/*	88,000	85,100	210	Ready	2,900	3.3%	419.0
/guides/*	39,000	37,600	140	Ready	1,400	3.6%	278.6
/tag/*	27,000	1,100	2,400	Ready	25,900	95.9%	11.3
/*?page=*	74,000	41,000	3,900	Ready	33,000	44.6%	19.0
/wp-json/*	15,800	0	1,200	Ready	15,800	100.0%	13.2
/cart/*	9,100	0	300	Ready	9,100	100.0%	30.3
/stores/*/events/*	21,000	3,400	5,300	Ready	17,600	83.8%	4.0

The Workspace tab with the worked example. Lime cells are inputs; blue cells are formulas.

Fill it with DataForSEO

This template is built to run on **DataForSEO On-Page API**: a crawl of your site: URLs, status codes, canonicals, inlinks and sitemap membership.

DataForSEO supplies the URL or query list and the audit trail; you add the measurements listed below.

1. Request the data

```
POST https://api.dataforseo.com/v3/on_page/task_post
[
  {
    "target": "yourdomain.com",
    "max_crawl_pages": 1000
  }
]
```

Post the crawl task, wait for it to finish, then request on_page/pages with {"id": "<task id>", "limit": 1000} and save that response.

2. Save the full JSON response

Save the complete response, not just the task ID, next to the template folder.

3. Convert it to the input CSV

The script in the Automation folder runs offline: no API call, no credentials.

```
python3 Automation/dataforseo_import.py import --template 14 --input response.json --output my-inputs.csv
```

Fills automatically	You add
Source reference (audit trail)	URL pattern, Verified bot requests, Requests to valuable URLs, Unique requested URLs, Average internal inlinks, Next action

Source reference fills automatically with the task ID, endpoint, location, language, device and snapshot time, so every row is traceable. Paste the finished CSV at Workspace cell A13. Estimate the cost first with the DataForSEO Cost Calculator (template 38).

Input dictionary

Keep the CSV columns in this order and paste values at Workspace cell A13. A missing observation stays blank; use 0 only for a measured zero or a verified No.

CSV field	Column	Meaning and rule	Typical source
`pattern`	URL pattern	Mutually exclusive URL pattern or template group.	Your records or manual review
`verified_requests`	Verified bot requests	Count from server logs after verifying the bot identity.	Your records or manual review
`valuable_requests`	Requests to valuable URLs	Requests inside the same group to approved canonical indexable URLs.	Your records or manual review
`unique_urls`	Unique requested URLs	Distinct URLs requested by the verified bot.	Your records or manual review
`inlinks`	Average internal inlinks	Crawl-derived average inlinks for this same group.	Your records or manual review
`action`	Next action	Document a proposed navigation, canonical, status or URL-generation change.	Your records or manual review
`source_ref`	Source reference	Task ID, export filename and observation scope. Retain the untouched source response separately. Optional.	DataForSEO: source_ref

The worked example

Every workbook ships filled with data from Northbeam Outdoor, a fictional outdoor-gear retailer. This one has 12 example rows.

12	0	665,800	1,145,900
Records	Rows with input issues	Requests to review	Verified requests in scope

Filter combinations and sort parameters account for most of the example's waste. Product URLs are 95% valuable and should be left alone.

Top 10 by requests to review

#	URL pattern	Requests to review
1	/*?color=*&size=*	231,200
2	/*?sort=*	177,900
3	/search?q=*	96,000
4	/*?sessionid=*	41,000
5	/*?page=*	33,000

Crawl Budget Analysis Template · Dashboard

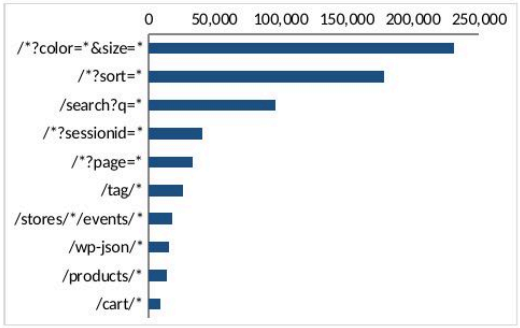
Everything on this sheet updates from the Workspace tab. Nothing here needs editing.

RECORDS	ROWS WITH INPUT ISSUES	REQUESTS TO REVIEW	VERIFIED REQUESTS IN SCOPE
12	0	665,800	1,145,900

TOP 10 BY REQUESTS TO REVIEW

Ranked automatically from complete rows in Workspace.

#	URL pattern	Requests to review
1	/*?color=*&size=*	231,200
2	/*?sort=*	177,900
3	/search?q=*	96,000
4	/*?sessionid=*	41,000
5	/*?page=*	33,000
6	/tag/*	25,900
7	/stores/*/events/*	17,600
8	/wp-json/*	15,800
9	/products/*	14,000
10	/cart/*	9,100



ACTION LIST: FIRST 15 ROWS THAT NEED ATTENTION

The Dashboard updates automatically from the Workspace tab.

Reading the result honestly

Redirect checks, robots fetching and discovery can be legitimate. The metric finds work to inspect.

Limits. On-Page is a site crawl, not Googlebot logs. Server request data must come from your verified logs.

Mistakes to avoid

Using unverified user agents. Fake Googlebots are common. Verify before analyzing.

Blocking parameters that carry canonical signals. If a parameter URL is canonicalized, blocking it in robots.txt stops Google from seeing the canonical. Choose the fix deliberately.

Treating crawl budget as a problem on small sites. Sites with a few thousand URLs rarely have crawl budget issues. Check indexation first.

Tips from the field

Cut internal links to wasteful patterns before blocking them; links create the demand.

Track the same patterns monthly to confirm fixes reduce waste.

Pair with the indexability audit to check what's valuable.

Questions people ask

What is crawl budget?

The number of URLs Googlebot can and wants to crawl on your site in a given time. It matters mostly for very large or frequently changing sites.

How do I analyze crawl budget with log files?

Filter server logs to verified Googlebot, group requests by URL pattern, and compare requests on valuable URLs with requests on parameters, duplicates and dead ends.

What causes crawl waste?

Faceted navigation, sort and filter parameters, session IDs, internal search pages, infinite calendars and expired content that keeps getting linked.

Should I block parameters in robots.txt?

Sometimes. It's effective for pure waste, but it hides canonicals and noindex from Google. For duplicates with signals, prefer removing links and using canonicals.

Templates that work well with this one

Technical SEO Prioritization Template
Turn a technical SEO audit into a ranked fix queue by recovered clicks, profit and profit per hour of dev time.

Indexability Audit Template
Check status, robots, noindex and canonical on every page you want indexed, then see exactly what to fix first.

XML Sitemap Audit Template
Reconcile sitemap URLs against status, canonical, robots and noindex to see what to add, remove or fix.

Internal Linking Opportunity Template
Score each source-to-target internal link by expected visits and contribution per hour, then verify it went live.

Programmatic SEO Checklist
QA every generated page for verified data, unique value, indexability, links and rendering before release.

Core Web Vitals Tracker
Log p75 LCP, INP and CLS by URL and device, test against Google's good thresholds and see what fails.

Read the full tutorial online: **optimizationtheory.com/download-view/crawl-budget-analysis-template/**
Get all six templates in the **Technical SEO Toolkit**, or all 40 in the **Complete SEO & GEO Toolkit**.