seogirl

Independent research

# Inside a Real Purge

A close look at a real purge, merge and keep sheet for a large, news-heavy information site. What drove nearly 1,900 decisions, and what the sheet never recorded.

Archive review · Site anonymized

NEARLY 1,900 URLS, ONE DECISION EACH




**~99%**
of purged pages were news

**4 in 10**
purged pages had backlinks

**Few**
had a recorded redirect target

**~1,900**
URLs reviewed

**~85%**
marked for purge

**6**
evidence gates compared

**Ameneh Saeednia**
Senior SEO and Generative Search Strategist

Independent research
seogirl.ca

About this study

# One spreadsheet, nearly 1,900 decisions

Most content-pruning advice is written in general terms. This is what a real decision sheet looked like, from a past project I reviewed.

1

### What it is

A **purge, merge and keep sheet** for one large, news-heavy information site. Every URL had traffic, search and link data, plus three decision columns: **SEO, commercial and final**.

2

### The goal

Find out **what actually drove the decisions**, and compare the sheet with the evidence gates in my *Content Review Decisions* playbook.

3

### What I looked at

Decisions by category and page age, link data on purged pages, agreement between review layers, and what the sheet recorded about **URL handling**.

4

### How it's reported

The site, niche and dates are withheld. Figures are **rounded shares**, and performance is described in general terms only.

### What the sheet contained

Analytics and Search Console exports, backlink data and a site crawl, all joined per URL. It's a typical, well-built starting point. This review is about **how it was used**, not about the site's results.

The short version

# What the sheet shows

| | |
|---|---|
| ~99% | **of purged URLs were news articles.** Outside news, only a handful of pages were cut. Commercial guides and reviews were almost untouched. |
| Small | Small share of recorded clicks and conversions. This does not establish that removal was low-risk. |
| 4 in 10 | **purged pages had at least one referring domain.** A few had ten or more. The most-linked one had **no search clicks at all**. |
| Age | **separated the decisions better than clicks did.** Every page from the current year was kept. Almost everything five or more years old was cut. |
| Every row | **where the SEO and commercial decisions matched.** The only changes came at the final stage, and all of them went from keep to purge. |

### What it means

The sheet emphasized old news with low recorded performance. That pattern alone does not validate deletion. Redirects, evidence of value, dependencies and implementation outcomes still require review.

Finding 1

# News was the whole story

The purge was really a **news archive cleanup**. Commercial pages were almost untouched.

| Page type | Decision pattern |
| --- | --- |
| **News articles** | About **9 in 10** purged |
| **Regional law and regulation pages** | About 9 in 10 **kept** |
| **Commercial guides and reviews** | Almost all **kept**, plus the only two merges |
| **Product and other pages** | Almost all **kept** |

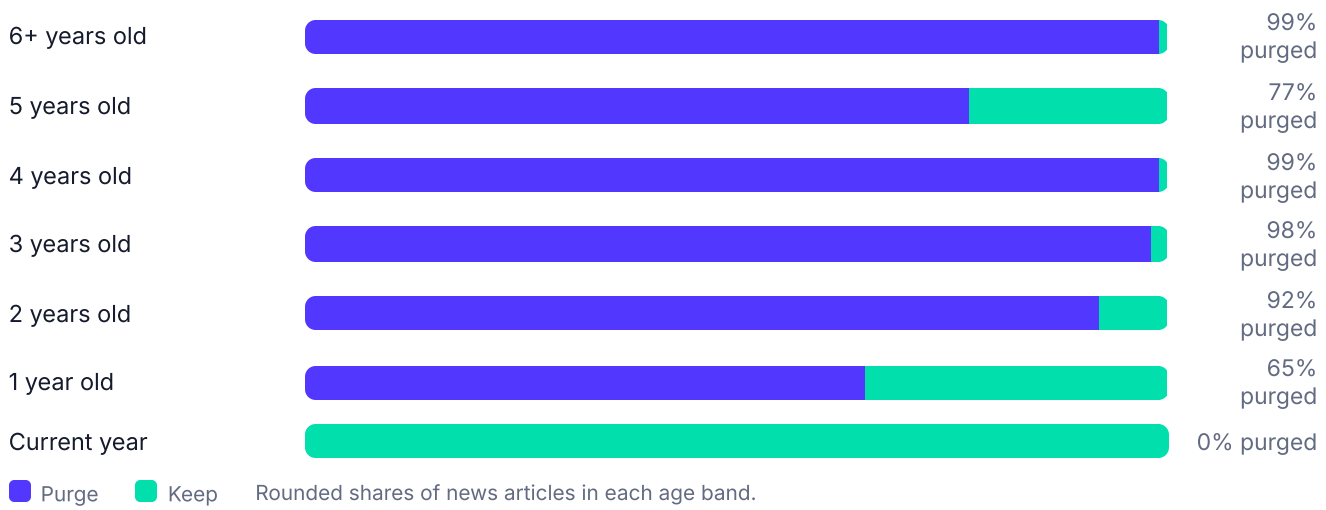
### Why that matters

News drew a meaningful share of visits but **almost no conversions**, while commercial guides drove most of them. That's the classic case for pruning, and it explains why **the commercial and SEO views never disagreed**: the cut didn't touch revenue pages.

Finding 2

# Age decided more than traffic

Clicks overlapped heavily between kept and purged pages. **Page age** didn't.




### Low traffic didn't mean purge

**More than half of the kept news articles** had almost no search clicks. Meanwhile some purged pages had dozens, and a few had hundreds. Traffic alone wouldn't reproduce these decisions.

### The likely rule, unwritten

Everything from the current year was kept, and almost everything older went. The sheet never states an age rule, but the pattern is clear. **A rule that drives the outcome should be written down**, with its exceptions.

Finding 3

# The links left with the pages

Purged pages weren't link-free, but the sheet recorded a destination for **almost none** of them.

| Referring domains on purged pages | Share of purged pages |
|---|---|
| **None** | About 58% |
| **1 to 2** | About 35% |
| **3 to 9** | About 7% |
| **10 or more** | **Under 1%** |

### What the sheet tracked

A "merge with this URL" column existed but was **empty for both merges**. Only a few notes in the whole sheet mention a redirect or merge target. For nearly all purged pages, **no destination is recorded**.

### What I'd add

For any purged page with links, record **redirect to a relevant page, or 410**, and the target URL. A news story with dozens of referring domains and no clicks may still be worth redirecting to a relevant hub, if one exists.

Referring-domain figures are per page, from a third-party backlink tool. The same domain can link to several pages, so these aren't unique domain counts.

Finding 4

# Three review layers, one answer

The sheet had separate SEO, commercial and final columns. On **every row**, the SEO and commercial decisions were identical.

| Stage | What happened |
| --- | --- |
| **SEO decision** | Keep, purge or merge for every URL |
| **Commercial feedback** | **Identical** to the SEO decision on every row |
| **Final action** | About 1% of pages moved from keep to purge. None moved the other way |

### The only changes

The pages that moved at the final stage were **low-traffic news articles**, each with about one search click.

### Design lesson

Perfect agreement can mean genuine alignment, or a second review that started from the first answer. Either way, it **adds little independent evidence**. A **blind review** of a small sample is an easy way to test the decisions.

What it means

# Mapping the sheet to the six evidence gates

How this real sheet compares with the gates in my *Content Review Decisions* playbook.

| Gate | In the sheet? | What was there, or missing |
|---|---|---|
| **Purpose** | Partly | Category labels only. No reader job or owner per page. |
| **Access** | Partly | A crawl export, but no recorded status, canonical or index decision. |
| **Value** | No | No accuracy or originality note. Old news was treated as a group. |
| **Overlap** | No | Two merges, with no recorded target. |
| **Performance** | **Yes** | Sessions, goals, clicks, impressions, position. The strongest part of the sheet. |
| **Dependencies** | Partly | Backlinks and referring domains collected, but no redirect plan. |

### The takeaway

A good purge sheet is mostly built. This one had the data to decide **what to cut**. What it lacked was the record of **where each URL goes next**, and a written rule for why. That's exactly the gap the playbook's record template is designed to close.

Limits and sources

# What this review can and can’t say

### No outcome data

The sheet doesn't record whether the purge was implemented, how URLs were handled, or what happened to traffic afterwards. **No result is claimed.**

### Rounded and generalized

The site, niche, dates and absolute figures are withheld. Shares are rounded. Performance is described in general terms only.

### One site, one moment

A single large, news-heavy information site. It isn't representative of e-commerce, SaaS or publishers.

### Why publish it

Real pruning sheets are rarely shown. The

The source workbook is not published. This review reports rounded, aggregate observations to explain its decision process; row-level data and identifying URLs are withheld.

01 A purge, merge and keep sheet from a past project
The source sheet is not published. Rounded aggregate observations are reported here; source rows, identifying URLs and private performance totals are withheld.

02 **Content Review Decisions. SEO Girl**
The six evidence gates and the review record used for the comparison.
seogirl.ca/skills/content-review-decisions

03 **How HTTP status codes affect Google's crawlers. Google**
Background for redirect and 404/410 handling.
developers.google.com/crawling/docs/troubleshooting/http-status-codes

By Ameneh Saeednia, seogirl.ca, prepared with AI assistance. Site and niche anonymized; figures rounded.