# SEO Audit & Web Performance Dashboard using Power BI

## 1. Objective

The goal of this project was to analyze SEO crawl data from Screaming Frog (SEO Audit Application) to find and showcase indexability, crawl budget usage, thin content, canonical issues and linking structures.

**This project demonstrates:**

- Clean and prepare messy SEO crawl data.
- Create calculated columns & measures.
- Build insightful dashboards with multiple perspectives of SEO health.

## 2. Dataset & Cleaning

**Dataset Information:**

- **Source**: Exported from Screaming Frog SEO Spider (CSV format).
- **Target Site:** FCC Co. Ltd (https://www.fcc-na.com), a Japanese automotive parts supplier.
- The dataset is **Publicly Accessible**.

**Data Features:**

- URL addresses crawled (1,027 pages).
- Indexability status (indexable, no-index, blocked, redirected).
- HTTP status codes.
- Crawl depth (hierarchical level).
- Word count per page.
- Metadata (titles, descriptions, H1s).
- Content type (HTML, image, PDF, CSS, etc.).
- Internal linking (inlinks, % share of total).

**Cleaning Steps**:

- Removed empty columns.
- Renamed columns for consistency.
- No modeling or multiple tables used, only one flat table (SEO_Audit).

All transformations beyond cleaning were handled with calculated columns and measures in the report view.

## 3. Calculated Columns & Measures

**Examples created in DAX:**

- **Url_Type**: Classifies URLs into “Indexable”, “No HTML”, “Blocked”, “Redirected”, etc.
- **Thin Content Cluster:** Buckets pages by word count (No Content, High Thin Content Risk, Normal URLs).
- **Inlinks Cluster:** Groups URLs by % of inlinks received.
- **Crawl Budget Type:** Buckets wasted, blocked, and crawled URLs.
- **KPI Cards:** Distinct counts of indexable, no-index, redirected, canonicalized, etc.

## 4. Dashboard Pages & Highlights

**Page 1: Macro Data**

- Overall indexability split (Indexable vs Non-Indexable).
- Crawl depth analysis (how deep crawled pages are distributed).
- Crawl budget waste vs useful crawls.
- URL Type and Content Type slicers for interactivity.

**Page 2: Web Performance & Metadata**

- Average response time across crawl depth.
- Missing or duplicate titles, descriptions, H1 tags.
- Metadata quality analysis with pie charts and duplicate tables.

**Page 3: Canonicals & No-Index Analysis**

- Canonical type distribution (canonicalized, duplicate canonicals, non-indexable).
- Lost links due to NoIndex or canonicalization.
- Pie charts for quick breakdown of non-indexable URLs.

**Page 4: Crawl Budget Analysis**

- Overall crawl budget utilization.
- % wasted crawls (blocked, redirected or no HTML pages).
- Visual showing wasted vs valid crawl gain by cluster.

**Page 5: Inlink Analysis**

- Distribution of internal link equity (inlinks).
- % contribution of links by cluster.
- Comparison of URL Types vs Inlink Weights.

## 5. Key Insights

**Indexability & Crawl Efficiency:**

- 78% of URLs are indexable and 21% are non-indexable or wasted (redirects, blocked, canonicalised).
- A significant portion of crawl budget is being consumed by redirected and blocked pages, reducing efficiency.

**Content Quality:**

- The Thin Content analysis shows a large amount of pages with very low word counts (≤100 or ≤300 words) making them weak candidates for ranking.
- Metadata checks shows some pages lack meta descriptions or short titles which can impacted click-through rates in search results.

**Technical Performance:**

- Response Time distribution shows most pages load in acceptable ranges (<1s) but there are clusters above 2 seconds which may hurt SEO performance and user experience.

**Site Structure & Internal Linking:**

- Crawl Depth analysis shows that a portion of content sits 4+ levels deep which lowering its visibility to crawlers.
- Inlinks distribution shows a little set of pages capture the majority of internal links but many others are weakly linked or isolated which may reducing their crawl and ranking potential.

## 6. Interactive Dashboard

- **View Interactive Dashboard**