



# Site Audit Sample

Version 1.0

BREAKDOWN OF ISSUES .....	5
1.0 Content .....	5
1.1 Issues .....	6
1.2 Recommendations .....	11
2.0 Crawling and Indexing .....	13
2.1 Issues .....	14
2.2 Recommendations .....	18
3.0 Sitemap .....	19
3.1 Issues .....	20
3.2 Recommendations .....	23
4.0 Robots.txt .....	24
4.1 Issues .....	25
4.2 Recommendations .....	26
5.1 Issues .....	28
5.2 Recommendations .....	30
6.0 Architecture .....	31
6.1 Issues .....	31
6.2 Recommendations .....	34
7.1 Issues .....	36
7.2 Recommendations .....	38
8.0 Inbound Links .....	40
8.1 Issues .....	41
8.2 Recommendations .....	44
9.0 Semantic Markup .....	45
9.1 Issues .....	46
10.0 Images .....	47
10.1 Issues .....	47

10.2 Recommendations .....	49
11.1 Issues.....	51
11.2 Recommendations .....	52
12.0 Titles .....	53
12.1 Issues.....	54
12.2 Recommendations .....	55
13.0 Meta Descriptions.....	56
13.1 Issues.....	57
13.2 Recommendations .....	58
14.0 Site Speed.....	59
14.1 Issues.....	60
14.2 Recommendations .....	62
15.2 Recommendations .....	64
16.0 Social .....	65
16.1 Issues.....	65
16.2 Recommendations .....	68

## EXECUTIVE SUMMARY

To easily navigate back to the Executive Summary, if you follow any of the links to pages deeper in this document, simply press Ctrl-G (Mac: Command-Option-G) and enter 13 to navigate back to the Executive Summary (which normally falls on page 13).

We provide the chart below as a quick view of the issues detected with SampleSite. The implementation priority is specified in ascending order, with 1 being the most important recommendation on down. (list out ranking 1-5 and what does it mean for each rank? Why rank it 1 instead of 2?)

Recommendations	Content Support	Technical Support	Priority
Content	Yes	Yes	1
Crawling and Indexing	No	Yes	1
Sitemap	No	Yes	1
Robots.txt	No	Yes	1
Canonicalization	No	Yes	1
Architecture	No	Yes	1
Internal Links	No	Yes	2
Inbound Links	No	Yes	1
Semantic Markup	No	Yes	3
Images	No	Yes	3
Videos	Yes	Yes	3
Titles	Yes	Yes	3
Meta Descriptions	Yes	Yes	3
Site Speed	No	Yes	1
Mobile	No	Yes	3
Social	Yes	Yes	2

## BREAKDOWN OF ISSUES

### 1.0 Content

Content Issues	
Implementation Priority	1
Content Support	Yes
Technical Support	Yes

Since the beginning, search engines have sought to serve up pages that searchers like and engage with, without having to immediately return to the search results page to try again. Taken in aggregate over millions of queries a day, search engines build up a good pool of data to judge the quality of their results.

Google, in particular, uses data from Chrome, its toolbar, and Android devices to measure these engagement metrics. Authority, expertise, and trustworthiness are key components of what Google considers quality content. The content also must be engaging and also must closely match user intent. Freshness and uniqueness of content also matters. Content should also be written at a reading level that matches the reading level of your target audience. It's therefore essential that each page on the site includes carefully crafted, high quality content that is strategically optimized to rank for a site's targeted keywords.

Because there are so many sites with similar content, it's important to stand out from your competitors. Top-notch content can help greatly with this. Plus, having well-written, relevant content provides a great user experience and can enhance conversions.

Search engines are getting smarter. Creating excellent content is even more important in the era of Google's RankBrain. RankBrain has been part of Google's algorithm since late 2015. It relies on machine learning to match queries to content. RankBrain is the third most important signal in Google's algorithm. This has meant a radical shift in retrieval of results.

The upshot is that if you want to beat the competition in search results, creating and maintaining superior content using natural language is an absolute necessity.

## 1.1 Issues

### *Site contains “thin” pages*

The site contains a lot of indexed pages with very little useful content. This not only can create index bloat, and wastes crawl resources, but also cause Google to penalize the site.

An example of thin content type on the site is abandoned X type pages. See below: [screenshot here]

Here’s [another example](#) that was found on the site. It was seen in results when drilling down by Explore > Organizations > Virginia: [screenshot here]

[This example](#) shows test pages started by a caused. These shouldn’t be displaying. [screenshot here]

Thin content can present dire issues for web sites. One of the most serious is a Google penalty called “Panda”. This major algorithm change was introduced in February 2011 to prevent what Google considers thin or poor quality content from being displayed in the top organic search results. Panda has had many updates since 2011 and is now a part of Google’s core algorithm. This means it has become a permanent Google ranking signal.

While Panda is flagged by an algorithmic signal, human raters can also be involved. Google’s quality raters are tasked with reviewing content quality of specific pages. Anything that is considered low quality by enough raters can trigger a Panda penalty.

The effect of a Google Panda penalty can be devastating to a web site. It can impact your bottom line. If pages from your site are pushed low in search results, it means a lot less organic traffic flowing to your site, which in turn results in decreased revenue.

Another problem with a Panda penalty is that it can hit a site repeatedly. This makes it even more important to be vigilant about thin or low quality content.

**We believe that the site was affected by the Panda 4.0 algorithm update in 2014.** Organic search experienced a major drop at the end of May/beginning of June 2014. It appears that the site has not fully recovered from this penalty.

The chart below illustrates the dip in organic search around the end of May 2014:

#### **Organic search by sessions, March 15 - September 15 2014** [screenshot here]

We compare two 3 month periods from 2014 (pre- and post-Panda) to show the effect on traffic to landing pages:

#### **Organic search by sessions and landing pages April 1 - May 31 2014**

The site may have recovered a bit around November 2014. It experienced a boost in December 2014 with the # campaign. However, it doesn't seem to have fully recovered. It's also important to keep in mind that **unless the problems that caused the penalty in the first place are addressed, the site can be hit with another Panda penalty at any time.** [screenshot here]

#### **Organic search by sessions, September 16 - March 14 2014** [screenshot here]

To resolve this issues, "thin" pages should be removed or marked as "no follow, no index" and a process should be put in place to programmatically check content quality before allowing users to publish their pages. Additionally, users should not be allowed to create duplicate or "test pages". Ideally each page would go through a human review prior to approval, but we would recommend a programmatic approach first.

#### *Pages with content duplicated on other pages*

The site has a lot of duplicated content on its site. Duplicate content is problematic and can also trigger a Panda penalty. It causes confusion for search engines. They don't know which page that should be indexed, they won't know which one to display in search engine results, which could result in dilution of link juice across multiple pages with the same content. The end result of duplicate content could be a loss in ranking.

One of the issues causing duplicate content on the site is duplicate URLs. This happens when variations of the same URL all resolve but are not canonicalized to one URL. This leaves search engines having to guess which URL is the preferred one. Here is an example of duplicate URLs that lead to the same content:

<https://www.SampleSite.com/team/village>

<https://www.SampleSite.com/team/village/>

<https://www.SampleSite.com/team/village.html>

<https://www.SampleSite.com/team/village.htm>

<https://www.SampleSite.com/team/village.aspx>

Another thing that can cause duplicate content is duplicated content across subdomains. Below is an example of different URLs with the same content. We've included an extended example of a page duplicated across subdomains in the [audit-data.xlsx file](#).

<https://www.SampleSite.com/story/example/related.html?page=29>

<https://pta.SampleSite.com//story/ example /related.html?page=29>

<https://svgives.SampleSite.com/story/ example /related.html?page=29>

<https://.SampleSite.com/story/ example r/related.html?page=29>

Another issue is that there is content on the site that there is duplicated on external sites without permission. This is called "content scraping". It's done as an attempt to attract traffic to a site. Here is an example of content scraped from one of the key landing pages on the site:  
[screenshot here]

To resolve these issues, select one URL you wish to use as the canonical URL. Use 301 (not 302) redirects that point to the canonical URL. [screenshot here]

### *Under-optimized pages*

Sometimes, in our haste to populate our sites with content, pages can be under-optimized for the keyword(s) you want them to rank for. One unintended consequence of this is called keyword cannibalization. This occurs when you have multiple pages optimized for the same keyword, but the search engines can't determine which page they should use to send searchers to for that keyword.

Below are examples where several pages are competing for some of the site's top keywords:

[screenshot here]

**"Sample keyword"** [screenshot here]



## *Pages with missing H1 tags*

Search engines use the H1 tag to help interpret what a page is about. H1 tags can be helpful for both users to scan the content and also help search engines better understand the structure of a page.

Using a well-optimized H1 tag could also possibly boost rankings when targeted keywords are used. But avoid keyword stuffing of H1 tags. Keyword stuffing can cause search engines to consider the page as spammy.

There are about 1,000 missing H1 tags on news.SampleSite.com. Most of these are on older pages and on subdomains. Example: <http://news.SampleSite.com/2015/10/get-featured-on-SampleSite/>. It may or may not be worth adding these retroactively. You may want to review the pages and consider adding them to the most significant pages.

In addition to structuring and optimizing H1 tags properly, ensure that all additional headings are used appropriately and are structured in the proper order. Assistive technologies for people with disabilities rely on the proper use of H1 tags and other header tags.

## *Top pages have low click through rates*

A high click through rate (CTR) is a good indication that users are finding your content on search engine results pages and are enticed to click through to the page. If using ads, it also means that users are finding the ads relevant to their needs.

CTR is the number of clicks divided by the impression count. Impressions are defined as “links to your site a user saw on Google search results, even if the link was not scrolled into view. However, if a user views only page 1 and the link is on page 2, the impression is not counted. The count is [aggregated by site or page](#). [Google says](#) that “with infinitely scrolling pages, such as image search, the impression might require the item to be scrolled into view.” ()

Some pages on SampleSite.com are getting a high number of impressions on Google but have a low CTR. These include landing pages for targeted keywords. The chart below shows the pages from March 4 - June 4 2016 with the highest impressions. The highlighted areas show the low CTR for <https://www.SampleSite.com/story/new> and <https://www.SampleSite.com/p/-ideas>, two important landing pages. [screenshot here]

The [latest SEO research](#) seems to indicate that click through rate from search is one of the many ranking factors.

### *Landing pages are under-optimized*

Below we present a screenshot from [report from Google Analytics](#) for a six-month period from December 6-June 5, 2016. [screenshot here]

Note the spammy page is the top landing page.

Clearly there are missed opportunities here. One of the problems is the [lack of directory structure](#) as discussed in this report.

Additionally, there seems to be a lack of strategic keyword research and content optimization strategy based on what Google is currently rewarding in search.

Here's an example of what is being rewarded for "personal page". Note SampleSite's page which is ranking #10. The URL, and title don't address the intent behind "personal page", nor do they use that phrases. SampleSite would see more click through from search if that landing pages was optimized for the term for which it ranks. [screenshot here]

Here's another example where there seems to be a lack of keyword research and optimization strategy for top target terms. Here's a term that would make sense for SampleSite to own, but for which the site does not appear: [screenshot here]

"Sample keyword phrase" [screenshot here]

The good news is that there are keywords that SampleSite selected (based on those being used in your Adwords account) for which the site is currently ranking and competitors are not. You can see that ranking data in the associated spreadsheet.

### *Pages contain multiple H1s*

There are many pages on the site with multiple H1 header tags. Multiple H1 tags are allowed in HTML5 but they should be used only when it makes sense to use them to markup section content. Using H1s and other header tags for text style purposes should be avoided. On many pages in SampleSite.com, H1s are used to style links (sign up link, e.g.).

## *Conversion pages in index*

Conversion — or “thank you” — pages should not be in search engine indexes. If they are indexed and start receiving organic traffic, your site’s conversion metrics will be inaccurate.

Some of SampleSite.com’s conversion pages are being indexed by search engines. Several are receiving organic traffic. We found that the following pages are indexed and are getting organic traffic:

<https://www.SampleSite.com/checkout/thankyou>

<https://www.SampleSite.com/p/Claim-Org-Thankyou>

<http://news.SampleSite.com/subscribe-thank-you/>

<https://www.SampleSite.com/p/Address-Change-ThankYou>

<https://www.SampleSite.com/p/Address-Change-ThankYou>

<https://www.SampleSite.com/p/EFT-Thankyou>

<https://www.SampleSite.com/p/e-news-thank-you>

To resolve this these pages should be blocked from crawling in the robot.txt file.

## **1.2 Recommendations**

- Recovering from a Panda penalty is complex. It involves a deep and thorough analysis of the site to discover what exactly is causing the issues. Because the issues causing Panda can be complex and recovery is not a quick or simple process, we recommend a Panda focused audit to deeply analysis and create a recovery plan for SampleSite’s Panda penalty.
- As a part of that Panda audit, we recommend an audit of duplicate and scraped content on the SampleSite.com site. This would include a thorough review of duplicate content to include same-domain and subdomain duplication, other URL duplication canonicalization issues, and scraped content review and recommendations. For known scraped content, you can take action to get scraped content removed from search engine results by contacting [Google](#) and [Bing](#).
- As we pointed out earlier, content is crucial. If your site doesn’t contain superior content, you will not have a chance against your competitors. This is even more

important because of the SampleSite.com Panda penalty. Without high-quality content, it will be difficult from SampleSite to recover from the penalty and avoid additional penalties in the future. We can create a content strategy for SampleSite and provide both writers who create highly targeted, well-written, search engine friendly, natural language content AND provide training to in-house writers around how to create content that will attract and engage users and help prevent content-related Google penalties. We can also create a plan to identify and systematically weed thin and spammy pages from the SampleSite.com site.

- Remove H1 and other header tags that are used to style link text. Use H1 tags and other header tags for content sectioning purposes only. There should only be one H1 per page.
- Provide training resources for user generated content about how to create SEO friendly pages.
- Make sure that conversion pages are not indexed; and ensure that all users who are donating see a donation thank you page.

## 2.0 Crawling and Indexing

Crawl Issues	
Implementation Priority	1
Content Support	No
Technical Support	Yes

Search engines frequently crawl websites in order to determine which ones should be indexed in their massive databases. Search engine crawlers — also known as robots or spiders — collect pages they deem high quality.

Over time, search engines will continue to request previously downloaded pages to check for content updates. A certain amount of bandwidth is then allocated for the periodic reviews based on the pages' calculated relevancy. Every time a page is downloaded, bandwidth is used, and once a website's allocated bandwidth limit is reached, no more pages will be crawled until the next review. This is referred to as a site's crawl budget, and it works just like any other budget. But in the case of a search engine, when your site's crawl budget has been exhausted, the bot moves on.

Since there is a limited amount of allotted bandwidth, it is crucial to direct the crawlers to the content you want to be included in the search engine's index of web content and eliminate any unnecessary or duplicate content. Your site's *crème de la crème*, if you will.

To use your site's crawl budget effectively, it's important to investigate errors reported in Google's Search Console (formerly Google Webmaster Tools) to make sure that those pages render properly. If a page can't be fixed, due diligence must be applied to make sure the page doesn't negatively impact the rest of the site. This can be done several ways:

- 301 redirect these pages to their new URLs. Use a 302 if the redirect is truly temporary (e.g., a product that's out of stock).
- If you have a lot of missing pages, allow them to return a 4xx (e.g., 404, 410) status code but be sure to remove these pages from your [XML sitemap\(s\)](#).

- Remove any internal links pointing to these error pages.
- If any of your 4xx pages have received significant traffic or links, you need to make sure you 301 redirect these pages.

## 2.1 Issues

### *Dips in crawl rate*

Crawl rate is based on how many requests per second search bots make to your site when it is crawling it. If a site experiences a dip in crawl rate, that means it isn't being crawled efficiently or thoroughly. Dips in crawl rate can have a major impact on a site. These are among the approaches to improve crawl rate (many of which will be addressed in subsequent sections of this report):

1. Improve [page speed](#). There is often a correlation between the time spent downloading a page and the pages crawled per day. It makes sense. The less time it takes GoogleBot and BingBot to crawl pages, the more pages it can crawl and index.
2. Block pages and sections of the site that shouldn't be crawled by robots.txt. The fewer pages you tell search bots not to crawl, the more pages will be crawled and indexed. (Note: use robots.txt carefully and wisely. More information here: <https://support.google.com/webmasters/answer/6062608?hl=en>)
3. Make sure your sitemaps are without errors and include pages that you want search bots to know about.
4. Fix duplicate content issues.
5. Make sure infinite crawl pages are addressed.
6. Correct HTTP header responses are vital for crawl efficiency.

The chart below shows the average page load time for the site from March 4 - June 4 2016. Note the correlation between lower crawl rate and time spent downloading. [screenshot here]

**Snapshot of crawl from May 2-May 8, 2016** [screenshot here]

### *Index bloat*

According to Google Search Console, the SampleSite.com site currently has 894,491 pages in Google's index. Yet only 13,568 landing pages received organic traffic last month. That means that only about 1.5% of landing pages are receiving any traffic. Most of those pages just received one visit in the past month. Only 348 landing pages received at least 10 visits.

In a perfect world, every page that Google indexes will find valuable enough to send organic traffic to your site. We don't live in a perfect world, and most sites will never achieve a 1:1 ratio, but we'd like to get your site as close to a 1:1 ratio as possible.

We're seeing a lot of content appear in search results that shouldn't. Examples are pages of incomplete campaigns as well as campaigns that were completed years ago. Pages in the index that serve little purpose to users can cause index bloat and waste your crawl budget.  
[screenshot here]

**Pages indexed by Google, October 25, 2015 - May 25, 2016 (total pages=894,491)**

**Landing pages with organic traffic, May 4, 2016 - June 4, 2016** [screenshot here]

**Landing pages with at least 10 sessions, May 4, 2016 - June 4, 2016** [screenshot here]

Bing is not nearly as efficient, however. It has crawled 840,011 this month and indexed 352,252 pages.

**Pages crawled by Bing, May 3, 2015 - June 3, 2016 (total pages=352,616)** [screenshot here]

### *Custom 404 page*

One way to pull visitors back into the site that have landed on a broken page is to serve up a custom 404 page. A well-designed custom 404 page will not only tell your users they content no

longer exists, but it can also serve as an opportunity to provide other helpful content to encourage users to explore your site further. SampleSite currently has a custom 404 page:

[screenshot here]

The SampleSite custom 404 page could use some improvement in design, usability and engagement. Here are some suggestions Google has offered for creating an effective 404 page that can help keep visitors on your site and help them find the information they're looking for:

- In order to prevent 404 pages from being indexed by Google and other search engines, make sure that your web server returns an actual 404 HTTP status code when a missing page is requested.
- Tell visitors clearly that the page they're looking for can't be found. Use language that is friendly and inviting.
- Include 404 in the page title, so you can easily track these pages in Google Analytics (or whatever analytics service you're using).
- Add links to your most popular articles or posts, as well as a link to your site's homepage.
- Ensure that the URLs you want users to go to are directly linked to from your custom 404 page.
- Have a site search box somewhere on your custom 404 page. (More sophisticated visitors tend to search; less advanced visitors tend to browse.)
- Consider providing a way for users to report a broken link.

There are also redirect issues with this page:

- The homepage URL links to <https://www.SampleSite.com/>. This URL uses a 302 redirect to <https://www.SampleSite.com/home/>.
- The URL for the giving catalog (<https://www.SampleSite.com/catalog>) has two redirects. The page uses a 301 redirect to <https://www.SampleSite.com/search.html>. This page uses a 302 redirect to <https://www.SampleSite.com/search/?action=search&controller=search&format=html>.

## *Paginated URLs*

Pagination has always been a sticky problem for search engines. They can cause inefficiencies with the crawl and excessive duplicate content. They can also create problems with undesirable pages ranking for important terms when search engines don't pick your preferred URL.

Fundamentally, pagination is an interface problem. Long, scrolling pages can cause latency, which can, in turn, cause visitors to leave your site (or bounce). In an effort to remedy this



issue, Google released some recommendations for how to resolve pagination issues in March 2012.

One of their solutions is to create a view all page and place a canonical tag on all the paginated content that points to the view all page. Most webmasters find this solution laughable since very few people (other than search engines that want to index as few pages as possible) want to view all running shoes, for example, on one page.

A more reasonable solution, Googlers posited, was to take advantage of the HTML 4/5 link element `rel="next"` and `rel="prev"`. This creates a chain of references that tells the search engines when they see page 3 of a string of paginated content, "Page 2 is the previous page, and Page 4 is the next page."

The SampleSite.com uses infinite scroll on many of their pages with extended content. We discuss infinite scroll and pagination [in this section](#). However, there are some parts of the site that do not use infinite scroll but do use pagination. These pages seem to be pages that display related causes of a particular fundraiser. Example: [screenshot here]

<https://www.SampleSite.com/story/sample>. There are many variants of this page and others that are causing duplicate content issues.

Many of these pages are appearing in organic search. Below is an example: [screenshot here]

SampleSite has implemented this code on the `/related_causes/` pages. However, the code on "related causes" pages seem to be implemented incorrectly. The code for the pagination in these pages is `<a rel="next">` but it should be `<link rel="next"`, e.g.).

Also, the `rel="next"` and `rel="prev"` code is in the `<body>` section of the page. [Google recommends](#) putting `rel prev/next` elements in the `<head>` of the page, not the not body.

You can learn more about how to resolve pagination issues from the [Google site for developers](#). The Search Engine Land site also has an [excellent post on this topic](#).

### *Site uses nofollow tags ineffectively*

The site needs to be reviewed for pages that should not be followed. The "nofollow" attribute is used when you don't want any links on a page to be followed by searchbots. They can also be used for individual links on pages. The idea is to save the crawl budget for pages/links that should be followed and omit those for which following them serves no real purpose. These include but may not be limited to login pages, user registration pages, and contact us pages.

Also, there may be links to external sites that for some reason may not be desirable for the searchbots to follow. An example is to an external site that is linked to numerous times. An example are the links to abc. We've noticed these links appearing on pages served up to Google and not to users (we discuss this behavior in [another section](#)). Although this is a reputable organization, there is no need to pass so much link juice onto an external site.

### *Navigation uses JavaScript*

JavaScript can be very useful for a site. It can, however, also cause problems. Best practice is not to use JavaScript for display of navigation links and content. We discuss this issue in the [Links are served with JavaScript section](#) in the report.

This is what users see on some pages see JavaScript is disabled: [screenshot here]

Other pages on the site are completely blank with JavaScript disabled. Clearly this is not desirable for SampleSite.com.

## **2.2 Recommendations**

- Serve a 404 (not found) or 410 (gone) response code for pages that no longer serve a purpose and are creating index bloat.
- Address crawl rate issues outlined in this section.
- Create an improved custom 404 to aid with usability and site engagement.
- Review the use of nofollow tags.

### 3.0 Sitemap

Sitemap Issues	
Implementation Priority	1
Content Support	No
Technical Support	Yes

XML sitemaps are designed to help search engines discover and index your site's most valuable pages. Think of these pages as a site's *crème de la crème* content.

The sitemap consists of an XML file that lists all the URLs of the site that you want to compete in organic search. This XML file also contains ancillary information, such as when a page has been updated, its update frequency, and relative importance.

Google recommends creating separate sitemaps for different types of content: images, videos, and text, for example. Then it's important to update them each time new content is published with a [multi-format XML sitemap](#), [image](#), or [video](#) XML sitemap.

It's important to update the sitemap only with the original versions of the URLs. They should not redirect or return errors. In other words, every page in the sitemap should return a 200 status code. Anything other than a 200 status code is regarded by the search engines as "dirt" in a sitemap. Duane Forrester with Bing said that Bing only tolerates a 1% of dirt in a sitemap. When asked what qualified as dirt, he replied, "Anything that doesn't return a 200 status code."

Finally, you should upload these sitemaps to Google and Bing's webmaster tools. Google will tell you how many of the submitted URLs have been submitted and how many of those have been indexed, whereas Bing just tells you how many URLs were submitted.

## 3.1 Issues

### *Sitemaps not submitted to Google or Bing*

Google does a pretty good job at finding a site's content; however, Bing is much more dependent on XML sitemaps for discovery. Not submitting your XML sitemap(s) to the search engines — especially Bing — is like leaving money on the table.

No sitemap has been submitted to Bing. A sitemap has been submitted to Google but it has errors and warnings that need to be addressed. These errors are found in [Google Search Console > Crawl > Sitemaps](#). [screenshot here]

### *Errors in XML sitemap*

It's important to search engines to have “clean” sitemaps. That is, sitemaps without coding errors and that aren't too big.

Google Search Console reports 1 error and 24 warnings. See the **sitemaps** section of the Google Search Console tool for details. Also see the [Sitemap too large section](#) later in this report.

**Google sitemap error report** [screenshot here]

### *Non-200 pages in sitemap*

Sitemaps should contain only URLs that return a 200 (found) HTTP status code. Directing searchbots to broken links and redirected content confuses searchbots and results in lost opportunities to let the search engines know about the site structure.

We analyzed the 5 separate [sitemap files](#) to see if there were non-200 status codes. Our findings: **all the sitemaps contain non-200 response codes.**

[https://www.SampleSite.com/assets/sitemaps/SampleSite\\_1.xml](https://www.SampleSite.com/assets/sitemaps/SampleSite_1.xml): Every URL returned a non-200 status code.

- 4 URLs return a 301 status code
- 1 URL returns a 404 status code

- 49,996 URLs return a 302 status code

[https://www.SampleSite.com/assets/sitemaps/SampleSite\\_2.xml](https://www.SampleSite.com/assets/sitemaps/SampleSite_2.xml): Every URL returned a non-200 status code.

- 2 URLs return a 301 status code
- 7 URLs return a 404 status code
- 49,993 URLs return a 302 status code

[https://www.SampleSite.com/assets/sitemaps/SampleSite\\_3.xml](https://www.SampleSite.com/assets/sitemaps/SampleSite_3.xml): Every URL returned a non-200 status code.

- 6 URLs return a 301 status code
- 14 URLs return a 404 status code
- 49,980 URLs return a 302 status code

[https://www.SampleSite.com/assets/sitemaps/SampleSite\\_4.xml](https://www.SampleSite.com/assets/sitemaps/SampleSite_4.xml): A significant number of URLs that returned non-200 response were present.

- 5 URLs return a 301 status code
- 3,375 URLs return a 404 status code
- 25,770 URLs return a 302 status code
- 1 URL returns a 500 status code

[https://www.SampleSite.com/assets/sitemaps/SampleSite\\_5.xml](https://www.SampleSite.com/assets/sitemaps/SampleSite_5.xml):

- 6 URLs return a 301 status code
- 1,140 URLs return a 404 status code
- 13 URLs return a 302 status code

### *Non-canonical URLs in sitemap*

All URLs in a sitemap should use the canonical URL. Remember: You are telling the search engines “these are my favorite pages that I want searchers to find”.

There are many non-canonical URLs in the sitemaps. Here is an example from

[https://www.SampleSite.com/assets/sitemaps/SampleSite\\_4.xml](https://www.SampleSite.com/assets/sitemaps/SampleSite_4.xml)

<https://www.SampleSite.com/story/sample> redirects to  
<https://www.SampleSite.com/story/sample>.

### *Sitemap too large*

Google recommends that sitemaps have less than 50k URLs or are 10MB unzipped. Google Search console reports that \*.SampleSite\_1.xml is too big. But when we pulled the sitemap files on June 1, 2016, sitemaps 1-4 all had more than 50k URLs:

SampleSite\_1.xml: 50,008 URLs

SampleSite\_2.xml: 50,018 URLs

SampleSite\_3.xml: 50,011 URLs

SampleSite\_4.xml: 50,036 URLs

SampleSite\_5.xml: 15,129 URLs

### *Low rate of URLs in sitemap indexed*

In an ideal world, 100% of all URLs submitted in a sitemap would be indexed by search engines. Unfortunately, this isn't an ideal world and reaching this goal can be difficult. But it's important to have as many URLs in the sitemap as possible. A site can lose revenue if pages in the sitemap (which are presumably the ones that are most important) don't get indexed.

We did an analysis (view [Index Diagnostics report](#)) of the number of pages submitted in SampleSite.com sitemaps and compared this number to the actual number of pages that are indexed by Google. This is called the sitemap ratio. We also compared the number of URLs indexed by Google that are also in the sitemap to the total number of URLs indexed. This is called the index ratio. Here are the results:

Pages submitted in sitemap: **215,454**

Pages in sitemap indexed by Google: **95,260** [screenshot here]

Total pages indexed by Google: 894,491

Pages in sitemap indexed by Google: 95,260

**Diagnosis:** Low Sitemap Ratio, Low Index Ratio

There are steps you can take to optimize your sitemaps. Here is some [great advice](#) from AJ Kohn.

## 3.2 Recommendations

- Ensure clean, well-optimized sitemaps are submitted to major search engines, including Bing.
  - Remove all URLs that return anything other than an HTTP status code of 200
  - Ensure sitemaps are of an acceptable length for search engines (less than 50k URLs or are 10MB unzipped)
  - Include only canonical URLs
- Do not exceed 50k URLs or 10MB unzipped.
- Create and upload video xml sitemaps.

## 4.0 Robots.txt

Robots.txt Issues	
Implementation Priority	1
Content Support	No
Technical Support	Yes

The robots.txt file is a very important text file placed in a subdomain's root directory. This file provides search engines directives regarding the crawling of the site, from blocking the crawling of specific pages or directories to the entire site.

Given the potential for disaster a robots.txt file carries, great care should be taken to ensure it follows the [specifications laid out here](#). If you want to learn more about how to configure a robots.txt file, there was a [great interactive guide](#) published to the Moz site.

It's not recommended to use robots.txt to prevent the indexing of site content that has quality inbound links pointing to it because if the search engines can't access a page, they can't credit you for those links. Similarly, you shouldn't use robots.txt to block a page that contains valuable links to other pages on your site. If the bots can't access the page, they can't follow links on that blocked page, which means authority will not be passed to those linked pages.

Even if you choose to not use the robots.txt file to prevent the crawling of any of the site pages, it's still recommended to use the robots.txt to specify the location of the site's sitemap.xml file, as specified in the [sitemaps.org documentation](#). If the site uses multiple sitemaps, then the link should point to a [sitemap index](#) that includes links to all of the site's sitemaps.



## 4.1 Issues

### *Pages missing from robots.txt*

Are there pages or directories that should be included in your robots.txt file? Here are some examples of items that might be considered to include:

Disallow: /cgi-bin/

Disallow: /tmp/

Disallow: /profile/

Disallow: /images/ (where /images is the directory for your template images)

The robots.txt file for SampleSite.com should also block cart and conversion pages. You can find your goal completion pages by navigating to Conversions > Goals > Goal URLs in Google Analytics.

It is not recommended to block login pages. Many customers get to their login page by searching for it in Google (e.g., *bank of america login*).

Google Search Console reports 134 resources blocked. These are blocked by external resources (there is nothing blocked in SampleSite.com's robots.txt file). All of these have this format:

<https://www.SampleSite.com/orders/logo?m=115100&s=146331693556857&t=f>

### *Blocking CSS and JavaScript*

Google strongly suggests that webmasters not block CSS and JavaScript files from Google in robots.txt files. Currently there are no blocked CSS or JavaScript files but we are including this for your information.

Gary Illyes from Google [said](#) the easiest way to grant Googlebot access to all your CSS and JavaScript is to add the following line items to your robots.txt file:

```
User-Agent: Googlebot
Allow: .js
Allow: .css
```

## 4.2 Recommendations

- Adding robots.txt directives can inadvertently create major problems if not done carefully. As a part of the Panda audit, we will include all the pages that should be included in the robots.txt file. This will involve a review URL parameters, directories, and other assets on the site for possible inclusion in robots.txt.

## 5.0 Canonicalization

Canonicalization Issues	
Implementation Priority	1
Content Support	No
Technical Support	Yes

In a perfect SEO world, the content on your site will only be accessible from one URL. This URL is referred to as the canonical URL. Reality is usually quite far from the ideal most SEOs strive for. That said, when canonical issues are scaled out to impact the site en masse, the site can really suffer. Imagine going into a horse race with the choice between one strong horse or several weaker horses. Which would you choose? The canonical version of a page of content is that strong, healthy horse.

To illustrate how subtle canonical issues can be, the following URLs may seem like they are the same page. But for a search engine these are actually different pages:

<http://samplesite.com>

<http://www.samplesite.com>

<https://www.samplesite.com>

<http://samplesite.com/index.html>

<http://www.samplesite.com/index.html>

<https://www.samplesite.com/index.html>

<http://www.SampleSite.com/Index.html>

<http://www2.samplesite.com/>

<http://www.samplesite.com/?id=15423&src=bingsrc=bing+ad>

Even though all of these URLs show the same content and refer to the same page, technically they're all different.

This kind of duplication creates a host of issues for your site:

- Dilution of page authority caused by both internal and external inbound links pointing to several URL versions of the same page.
- Self-cannibalization in the search results pages, with different versions of the same page competing for the same queries.
- Inefficient crawling since search engines spend more time crawling the same content and leave less time to discover new content.

Because of these issues, the best solution is to 301 redirect non-canonical URLs to the canonical version of the page whenever possible. When a redirect isn't feasible, the next best solution is to add a [rel="canonical" tag](#) in each of the site pages, giving credit to its canonical URL.

## 5.1 Issues

### *Homepage not canonicalized properly*

The homepage has no canonical URL associated with it. At minimum, a canonical link element (also informally referred to as a canonical tag) should be placed on the homepage so that duplicate versions will point back to it. It would look like this:

```
<link rel="canonical" href="http://www.yoursite.com" />
```

This element should go in the <head> section of the page.

Non-canonical versions of your homepage should be 301 redirected to the canonical version of the homepage.

Note that the Moz page authority score of the homepage is 42 (1-100 scale). Page authority indicates how well a web page is likely to rank in Google's search results. A score of 42 is low for a homepage. This is at least partially due to the canonicalization and redirect issues on SampleSite.com.

### *Canonical tag not included on the site pages*

A self-referencing canonical element should go on every page of the site to shore up its canonical pages from duplicate URLs. Query parameters can be added to URLs (like when someone shares a page from your site using Buffer), and sometimes these parameter pages can get indexed by Google. There are many pages missing a canonical tag on SampleSite.

We've included a canonical errors report in the [audit-data spreadsheet](#).

### *URLs can be accessed through non-secure protocols*

Good news: most of the SampleSite.com site uses the secure HTTPS protocol. The subdomain news.SampleSite.com is using a non-secure protocol. This also means that Google Search Console is not tracking news.SampleSite.com and Google Analytics will have data collection issues as well. We would recommend migrating news.SampleSite.com to secure HTTPS.

### *Parameter URLs creating duplicate content*

Query parameters can wreck a site with wasted crawl budget (at best) and duplicate content (at worst). Sites with many query parameters can literally make search engines crawl millions of duplicate pages because of all of the permutations possible from their query parameters. This can cause what is known as "index bloat": many more URLs being crawled than indexed.

SampleSite.com has many URLs with parameters. Google is not reporting a problems handling them, but that means that they are not causing duplicate content.

Bing is different. It doesn't necessarily automatically know how to deal with parameter URLs. If you do want BingBot to crawl certain parameter URLs, you can go to Webmaster Tools, go to Configure My Site > Ignore URL Parameters.

Note: be very careful when blocking parameters. Blocking parameters can cause unintended results. Read more about [configuring parameters in Google Search Console](#) and in [Bing Webmaster Tools](#).

You can also block parameters in your [robots.txt file](#) but again, you must be very careful in doing so.

## Other canonical issues

Many URLs have different canonicals than the source page. These canonicals then use 302 redirects to redirect to another URL.

Example: Canonical URL <https://www.SampleSite.com/story/Help-Me-Help> uses a 302 redirect to <https://www.SampleSite.com/us/story/Help-Me-Help>

There are also canonicals from `*/us/story/*` to `*/story*` according to Screaming Frog canonical errors report (however, we could not find some of these canonical URLs in the source code or with canonical checking tools). Example: source URL <https://www.SampleSite.com/us/story/SA> has <https://www.SampleSite.com/story/SA> as the canonical. If these canonicals do indeed exist, it can cause confusion for search engines because <https://www.SampleSite.com/story/SA> has a 302 redirect to <https://www.SampleSite.com/us/story/SA>.

Canonicals URLs on [news.SampleSite.com](https://news.SampleSite.com) are inconsistent. Some point to the non-secure [www.SampleSite.com](http://www.SampleSite.com) URL (which 302 redirects to the secure URL). Example: <http://news.SampleSite.com/tag/farmers-market/> has a canonical to <http://news.SampleSite.com/tag/farmers-market/>

There are also issues with canonicals that use 301 redirects. There were only a few of these.

At minimum, a canonical link element (also informally referred to as a canonical tag) should be placed on each page so that duplicate versions will point back to it. It would look like this:

```
<link rel="canonical" href="http://www.yoursite.com" />
```

This element should go in the `<head>` section of the page.

Non-canonical versions of each page should be 301 redirected to the canonical version of the page.

## 5.2 Recommendations

- As a part of the Panda audit, we recommend that a full duplicate content analysis be conducted discover and resolve all the issues discussed above.

## 6.0 Architecture

Architecture Issues	
Implementation Priority	1
Content Support	No
Technical Support	Yes

A good site URL structure should align itself with several best practices:

- Group topics together into category and subcategory pages to help both visitors and search engines to maintain topical relevance. Think of how a library is organized.
- Reflect the site's organization through levels of directories and subdirectories. If you have the misfortune of a totally flat architecture (i.e., all pages saved in the root), you'll need to get creative in how you send signals to search engines that certain pages are related. This can be accomplished, for example, by using breadcrumbs or adding a link to the category on the page.
- Use only lowercase to avoid confusion with users and search engines.
- Use dashes (-) to separate keywords in the URL.

## 6.1 Issues

### *Issues with directory structure*

Context is a ranking factor that can elevate your website past others competing for similar or related keywords, build topical domain authority, and increase conversions by having a higher number of pages occupying a larger percentage of market share for related keywords and queries. The architecture (and internal links) communicate context from page to page. The better structured this is, the better it is for users and search engines.

The [Content Drilldown report in Google Analytics](#) provides a good overview of a site's architecture. In a perfect world, a site will only have five to ten line items in this report because, like a library, your site should have top categories that spill into sub-categories, which spill into sub-sub-categories. This is one way to consolidate topical relevance.

The drilldown report shows that there are 1,358 level one page paths. This shows that directories on SampleSite.com are not optimized. It is recommended that SampleSite create a meaningful topical directory structure that helps users find what they need and gives search engines a better idea of how your site is structured.

It's also recommended that SampleSite consider changing their URLs to reflect hierarchical structure using targeted keywords. However, this would be done as a later phase after the directory structure is cleaned up.

### *Non-hierarchically organized URLs*

SampleSite.com's site architecture is rather convoluted from what can be gleaned from the Google Analytics Content Drilldown report. This report shows that most of the site's traffic flows from the /us/ directory. The SampleSite.com site has 81,046 line items in the the /us/ directory. This means the site has lots of content saved in the root directory. Some of these may be caused by query parameters, which can be cleaned up. You can read a blog post on how to do that [here](#).

The /us/ directory contains subdirectories that seem to be created for individual campaigns. Many of these subdirectories have no content and return a 404 code. This can hurt the site in search because search engines are dependent on a site's structure to decide between things like sending a searcher to an article/product page (for longer-tailed queries that searchers tend to run further into the sales cycle) or a category page (for head terms that searchers tend to use earlier in the sales cycle, for example).

One of the issues with SampleSite.com's site structure /us/ directory is duplicate directory names. There are two /us/home/ directories, two /us/search/ subdirectories, and two /us/about subdirectories. Also, there is yet another subdirectory named "story" in the /us/ subdirectory: /us/story/story/.

To complicate matters further, there is also a /story/ directory under /us/: /us/story/ and a /story/ directory that that exists outside of /us/: /story/.



We understand that restructuring SampleSite.com's site architecture can be complicated, but we would recommend that new landing pages be created for each type and category and resolving the URLs underneath those appropriate categories.

While we strongly recommend a structural overhaul, if this isn't possible, at the minimum we recommend creating breadcrumbs and enhancing the internal linking structure. We recommend that the [internal linking](#) structure be part of the additional Panda audit.

Breadcrumbs help orient users which in turn helps with usability (and usability helps with engagement). They also give search engines clues about site structure. This helps bots better to crawl a site. Google uses breadcrumb markup to categorize the information from a page in search results. See <https://developers.google.com/search/docs/data-types/breadcrumbs>

Google recommends using schema.org markup for breadcrumbs when possible to help Google and other search engines better "understand" site navigation. We also suggest that directories that are no longer valid be redirected using 301 redirects to the appropriate directory/page.

### *Site has separate mobile pages*

There doesn't appear to be a separate mobile site although there seems to be a different mobile view. Example: <https://www.SampleSite.com/mobile> redirects to <https://www.SampleSite.com/us/search/?action=search&controller=mobile&format=html>. The mobile view can be drilled down by category and state. Example: <https://www.SampleSite.com/us/search/?action=search&controller=mobile&format=html&rg=75> (Emergencies category). This could be creating duplicate content.

In the mobile view, clicking on the Explore link on the page returns users to the responsive (non-mobile) view. However, the two views of the Explore page (mobile and responsive) are different on a mobile device. See screenshots below.

#### **View of**

<https://www.SampleSite.com/us/search/?action=search&controller=mobile&format=html> in iPhone 5 device [screenshot here]

**View of** <https://www.SampleSite.com/search/explore> in iPhone 5 device [screenshot here]

Serving up different views of the same page using different user agents could be considered these URLs could cause cloaking issues. We discuss cloaking [in a separate section](#).

Since these URLs are no longer used, they should be redirected using 301 redirects to the appropriate responsive pages.

You can read more about mobile issues in the [Mobile section of the audit](#).

### *Infinite scroll pages aren't SEO friendly*

To make sure that search engines can crawl individual items linked from an infinite scroll page, make sure that you or your content management system produces a paginated series (component pages) to go along with your infinite scroll. Learn more about Google's recommendations [here](#). BuiltVisible also published a [great post](#) on how to make infinite scroll pages search friendly. Also check out the [demo by John Mueller](#) of Google to see one way of implementing pagination with infinite scroll.

### *Non-descriptive URLs*

The site uses many URLs with parameters. Example:

<https://www.SampleSite.com/us/search/?rg=5;30;31;32>. While these parameters don't seem to be causing any issues for Google, they don't give search engines (or users for that matter) any indication of the purpose of the URL. They also don't contain keywords (although it is never a good idea to create overly keyword-laden URLs because search engines frown on spammy URLs). This also creates issues for Google Analytics site search tracking. That said, with the other issues the site is facing, we don't recommend changing URLs at this point.

### *Broken links throughout the site*

Broken links are a sad experience for users. They want to read something you've linked to only to be disappointed when they see a 404 (not found) page. We added a spreadsheet with all broken links to external sites to the [audit-data spreadsheet](#).

## **6.2 Recommendations**

- Clean up directory structure. Create category and type landing pages.

- Implement breadcrumbs.
- Address infinite scrolling and pagination.
- Redirect old mobile URLs to their responsive equivalents.
- Deal with broken links. Are there pages that are broken that shouldn't be? Could content on some pages be revitalized or updated and served up to users?

## 7.0 Internal Links

Internal Link Issues	
Implementation Priority	2
Content Support	No
Technical Support	Yes

Internal links are not only used by visitors to browse through the site but also by search engines to discover content and pass authority to other internal pages.

To increase crawl efficiency and ensure that authority flows naturally from category pages down through the rest of the site and back up to the category pages, a few measures should be taken:

- Link directly to the [canonical URL](#). By linking to non-canonical URLs, you're wasting your precious crawl budget by sending search engine spiders on a scavenger hunt to discover content you want indexed by search engines.
- Use relevant, descriptive anchor text in the links without being unnecessarily repetitive or spammy.

## 7.1 Issues

### *Site doesn't link to pages with highest authority*

We found that pages on the SampleSite.com site don't often link to pages with the highest authority. However, we were unable to obtain an accurate picture of internal linking to the most authoritative pages in the site due to the questionable quality of the data we reviewed. For example, the third page listed in the top pages report we generated is <http://SampleSite.com/team/Dinners>. We are confident that SampleSite wouldn't consider this to be a top page. We've included the top pages report in the [audit-data spreadsheet](#).

## *Pages in the site pointing to 404 pages*

Our crawl of SampleSite.com shows that there are many pages in SampleSite.com pointing to pages that return 404 response codes. However, we are not confident of this data. For example, the homepage retrieved by a browser agent returns a 200 code but when viewed with the Googlebot agent it returns a 404 code. This means that Google can't "see" many 200 codes and returns them as 404 codes instead. We outline this issue in the [AngularJS](#) section. We recommend fixing the issues discussed in that section and then we can review actual broken links when the data is more reliable.

## *Daisy-chain URLs*

Even though Googler John Mueller [said in his October 15, 2015 hangout](#) that Google will follow up to five redirects, excessive redirects can cause latency and slow page load issues. We recommend you address redirects as much as possible, especially those that are causing infinite loops.

Unfortunately, sometimes it's hard to keep track of all the redirects on a site. This can result in unintentional redirects that loop back to each other (redirect loop) or loop back to themselves (infinite loop). We've included a tab in the [audit-data spreadsheet](#) that shows a list of redirect loops on the site and also discuss this in the [cloaking section](#) of this audit.

Note: This is not an exhaustive list. It's impossible to detect all of them due to the way the site is configured.

The homepage has several redirect hops. One of them includes a 302 redirect. Historically, it was considered bad practice to redirect using 302s. This is because they did not pass link signals to the URL being redirected to. [That is changing, however](#). Still, it's a good idea to avoid using them because 302 redirects mean "moved temporarily" whereas 301 redirects indicate "moved permanently". If you know that the URL you're redirecting from is never coming back, it's best to use a 301.

Here is the redirect path for the SampleSite homepage:

```
http://SampleSite.com (301) --> http://www.SampleSite.com (301) -->  
https://www.SampleSite.com (302) --> https://www.SampleSite.com/us/home/
```

Although Google can crawl these redirects, since the site is experiencing slow page loads (more on this later), we recommend removing the hops and using 301 redirects so that the point directly to the canonical version of the homepage.

We also noticed that a *SampleSite.org* domain exists. This redirects to the homepage at <https://www.SampleSite.com/home/> using the same 3 redirect hops as SampleSite.com does. A WHOIS lookup shows that the [SampleSite.org](https://www.SampleSite.org) domain expired in 2014. It's currently registered by Legatum in Dubai. It is advised that SampleSite implement a redirect to the canonical URL <https://www.SampleSite.com/home/> using a 301 redirect as described for SampleSite.com if this a SampleSite property. If it isn't, this should be reported to Google and Bing to disavow the site.

As we mentioned before, there are redirect issues on SampleSite.com that are causing infinite redirect loops.

Here is an example of an infinite loop on SampleSite.com category pages:

from <https://www.SampleSite.com/us/search/?orgScope=on&page=1>--> 302 redirect to  
<https://www.SampleSite.com/us/search/?action=search&controller=search&format=html&orgScope=on&page1>-> 302 redirect to  
<https://www.SampleSite.com/us/search/?action=search&controller=search&format=html&orgScope=on&page1> → redirect to  
<https://www.SampleSite.com/us/search/?action=search&controller=search&format=html&orgScope=on&page1> ... until Google gives up.

There are also infinite redirects from [mobile URLs](#). These result in infinite 302 redirects.

Example:

<https://www.SampleSite.com/us/search/?action=search&controller=mobile&format=html>

It's important to fix infinite loops so that search bots don't get hopelessly confused when trying to crawl your site. We discuss this further in the [AngularJS section](#) of the report.

## 7.2 Recommendations

- Because of the presence of questionable data, we suggest analyzing internal linking to high authority pages in a separate audit after SampleSite makes suggested content and

technical changes. This would also include creating a solid linking structure/strategy and reviewing internal anchor linking.

- Fix redirect loops and other redirect issues.

## 8.0 Inbound Links

Inbound Link Issues	
Implementation Priority	1
Content Support	No
Technical Support	Yes

One of the single most important elements of the Google ranking algorithm is the quantity and quality of inbound links (in other words, links from other sites to yours). Google also looks at the anchor text, surrounding text, alt text in images, page topic, and linking site topic as signals that determine what keywords a page should rank for.

There are a number of qualities search engine algorithms key in on:

*Quantity:* The total number of links coming in.

*Authority:* How high is the authority of the sites linking to your site, and where are the links coming from? Links from boilerplate areas of the site, like the footer, are considered lower quality than editorial links, as [WordPress learned the hard way](#).

*Topical Relevancy:* How relevant is the page linking to yours?

*Anchor Text:* The text of the link pointing to you. Search engines also look at the text immediately surrounding a link for topical clues.

*Velocity:* The rate at which new links are created and/or discovered. If your site creates content that goes viral or is highly relevant to current events, spikes are to be expected. But those spikes better be validated with other signals, such as social shares. Outside of virality, consistency (and staying below the radar) is key.



## 8.1 Issues

### *Decline in inbound links*

There appears to be a major decline in inbound links starting about December 2015. There could be a number of reasons for this. One of them could be due to a site redesign, Google finding 404 errors when trying to access pages on SampleSite.com that are being linked to. We will discuss this further in this section.

Another possibility is that there isn't enough compelling [content](#) to link to or calls to action to attract links. Site architecture, redirects, and other technical issues could also be factors.  
[screenshot here]

Generally, SampleSite.com has more inbound links than its competitors: [screenshot here]

### *Site contains spammy links and content*

#### **User-generated spam**

Google has detected user-generated spam on the SampleSite.com site. We investigated and our findings show that there is spam in user-generated campaign pages. Example:  
<https://www.SampleSite.com/team/Watch-Online> [screenshot here]

As a result, Google has applied a **manual spam action** to a portion of the site. Manual actions occur when a human on Google's team has determined that there is significant spammy content on a site. The manual actions on SampleSite.com are listed in Google Search Console under Search Traffic > Manual Actions > Partial matches.

This is a serious penalty and one that can be difficult to recover from. There is unfortunately no quick fix for a manual action because there is no guarantee that Google has caught all the spammy content. Because SampleSite.com has a high domain authority, it makes spam easier to rank which makes SampleSite a target for spammers who are always looking for sites to spam. It is therefore not out of the question that additional manual actions could be applied unless spam on SampleSite.com is eradicated.

User-generated spam is difficult to monitor and control. We recommend that a thorough review of spammy content be done so it can be removed and a subsequent reconsideration request be made to Google. Users who have created spammy content should be blocked

although it is always possible they can create new accounts. A link to report abusive, spammy, or inappropriate content on the site can also be added to fundraiser pages as a way to get the SampleSite community to help.

SampleSite should have a programmatic method to detect and prevent spam. Because this isn't always foolproof, however, we recommend that someone on SampleSite's team manually monitor spam on the site. This can be a huge task and requires someone savvy about the sneaky ways that spammers use to get their content placed on a site. We can train someone on SampleSite's staff. Alternatively, we can provide spam monitor services so spam can be quickly caught and removed from the site.

### **Unnatural backlinking**

We are also seeing what may be **unnatural backlinking** to SampleSite.com. Example: One of the sites with the most inbound links is *lifix.in*. Google Search Console reports there are 180,248 inbound links from this site. As you can see in the screenshot below, this site mirrors SampleSite.com: [screenshot here]

The practice of obtaining backlinks from a mirror site and any other unnatural linking methods is considered black hat and runs the risk of an unnatural link penalty or a Google Penguin penalty. Like Panda and manual action penalties, these can be a challenge to recover from. It is strongly recommended to take steps to identify disavow links from this site and any other risky backlinks. Additionally, if this site is scraping content from SampleSite, it should be reported to [Google](#) and [Bing](#).

### **Best linkable assets**

The goal of linkable assets should be to convert. A linkable asset should also be sharable. An example of a linkable asset on SampleSite.com is the [X page](#). This page would be a great choice for targeting as a conversion goal. However, there are no internal links on the page to other pages on the site except for footer links. There is also no text on the page to create calls to action.

Another linkable asset is the homepage. The homepage should have a prominent link above the fold for not just starting a fundraiser but also donating to one. Indiegogo has a "start a campaign" button in the upper left corner of every page: [screenshot here]

SampleSite has a link to start a fundraiser, but it is not distinguished in any way and therefore can be easily overlooked. We suggest making this element more visible and also including a button or link for a donation call to action. This would also use a bold distinguishable style and be placed above the fold. [screenshot here]

Further, the use of the term “Explore” doesn’t tell users where the link goes. What exactly will they be exploring when they click the link? Consider rethinking the use of the term.

### *Linkable assets aren’t top organic landing pages*

Using the example above, the giving card page is not a top landing page. This and other linkable assets on the site should be optimized.

### *Links to non-canonical URLs*

Don’t link internally to non-canonical pages. Sometimes this happens because someone will set up a link to non-www version of a page when the canonical page is the www version. Or they’ll link to the non-secure (http) page when the canonical version is the secure version (https). We’ve included a tab of links to non-canonical URLs in the [audit-data.xlsx workbook](#).

### *Inbound link anchor text*

We analyzed the top anchor phrases and terms for inbound links on SampleSite.com. The screenshots below show that while a few phrases (such as “donate now”) that are relevant, much of the most frequently used anchor text is not at all relevant to the site. These include “here” and “click here”. This also creates a poor user experience, particular for users with visual disabilities. They won’t be able to determine where “here” leads to if there is no context. Considering that up to 20% of the U.S. public has some sort of disability, web accessibility is also something to keep in mind.

It’s difficult to control the anchor text that people use to link to you but there are some things you can do to help influence what is linked to. Using keyword-rich text (but not spammy) on your site can help. You may also consider contacting people who link to you and suggesting your preferred anchor text. [screenshot here]

## 8.2 Recommendations

- Penalties and actions caused by user generated content, unnatural inbound links, and unnatural links warrant not just a reconsideration request, but a change in overall strategy and perhaps even your business model. We recommend removing any mirror sites and evaluate other inbound linking domains for unnatural links. A thorough review of the site should be conducted to fully identify spammy content and links, and strategies created to fix the issues and prevent them from reoccurring. This would be done as part of the additional audit we're proposing. We also suggest developing a method to programmatically detect and flag pages for suspicious content. We could train someone in-house to monitor the site to prevent user generated spam or we could provide a spam monitoring service as an adjunct to programmatic monitoring. Once the site is spam-free we can create a plan to submit a reconsideration request to Google to remove the penalty.
- Assess and investigate trends in inbound linking after issues described above are addressed.
- Leverage linkable assets.

## 9.0 Semantic Markup

Markup Issues	
Implementation Priority	3
Content Support	No
Technical Support	Yes

Using [schema.org](https://schema.org) with microdata in a site's HTML to declare specific elements allows search engines to better understand the meaning of its content and provide it with higher visibility in the search results through rich snippets. A rich snippet is essentially a bite-sized summary of the content that a user can expect to see on a page.

These snippets come in different forms: an image of a video screenshot, information about a concert, star ratings, cooking time for a recipe, etc. But all serve the purpose of providing sneak peek of the content on your site. It has also shown to [increase CTR from search](#).

In general, it's recommended to start using the following microdata when specifying this type of content in HTML:

- [Address](#)
- [Breadcrumbs](#)
- [Authors](#)
- [Events](#)
- [Product reviews](#)
- [Product information](#)
- [Ratings](#)
- [Videos](#)

This content can be validated through the [Structured Data Testing Tool](#) inside Google Search Console.

## 9.1 Issues

### *Insufficient structured data*

The SampleSite.com site only makes \*some\* use of the Open Graph Protocol (news.SampleSite.com). This allows content to be shared as rich objects across social media sites such as Facebook. There is also hAtom and hCard markup, mainly on /team/ pages. We discuss this more in the [Social section](#) of the report. We would recommend implementing the following schemas: event, video, organization, and payment service.

### *Errors in Structured Data*

Google Search Console reports errors for about half of the structured data markup on the site. We recommend either removing the markup altogether if it isn't serving a purpose or fix the issues. See screenshot below. [screenshot here]

## 9.2 Recommendations

- We suggest that SampleSite investigates using markup for a [sitelinks search box](#) and also if decides to use [breadcrumbs](#) on the site.
- Add additional structured data where appropriate. [screenshot here]
- Fix existing markup errors. [screenshot here]

## 10.0 Images

Image Issues	
Implementation Priority	3
Content Support	No
Technical Support	Yes

Images offer a great opportunity to earn visibility and additional organic traffic with image and universal search results.

The use of images on the site is also helpful to make pages more attractive, illustrate an idea, and attract visitors' attention.

Below are a few best practices in optimizing images for search:

- Include ALT text using specifically relevant keywords for each image without keyword stuffing them. Although ALT text is useful for SEO purposes, it is used primarily for the benefit of visitors with disabilities using assistive technologies. Learn more about this aspect of [ALT text](#).
- Optimize images file names so they include relevant keywords.

Read more about [Google's recommendations for image optimization](#).

### 10.1 Issues

#### *Images are on a different subdomain*

Many images on SampleSite.com are hosted on a subdomain, cloudfront.net. Rand Fishkin, founder of the SEO think tank [Moz](#), did a seminar on image search strategies where he specifically talked about the importance of keeping images on the same subdomain as the content you want to have ranked. He talks in the seminar about how Facebook developers saw

a reduction in visits from image search after moving their images to a subdomain. (You can [buy the seminar here](#), if you're interested.)

If you have load balance issues, you can load balance behind the firewall and have a separate server power your images. Whatever it takes on the backend, just have them show up close to the pages you want to rank in image search.

To measure the impact of not having images hosted on your main subdomain, we did a search in Google Images for *online*. SampleSite.com didn't appear on the first pages of search results for this query.

### *Large images*

With SEO being inextricably tied to user experience, search engines are trying to motivate webmasters to make their sites easier to use by adding page load time as a ranking factor. (Learn more in the [Site Speed section of the audit](#).)

One factor that can bloat page load times is large images. In the rush to publish, it's easy to overlook image size. Care should be taken to make sure images are under 100KB (and even smaller if you have an entire page of images).

The images in the carousel on the SampleSite homepage are too large. This not only affects loading time but also the user experience. When the images are above the fold and are purely decorative, they waste important real estate and therefore have a negative effect on both the user experience and conversions. It's always important to keep in mind that users do not want to scroll much. If your actionable content is located below decorative images, there's a good chance that users won't find it.

We've included a report of images over 100kb in the [audit-data spreadsheet](#). These are from the news.SampleSite.com blog site. Unfortunately, this report does not include images on SampleSite.com because of the external location of images on the site.

### *Images with no or irrelevant ALT text*

There are a lot of images on the SampleSite.com site that do not have alt text. Each image should have descriptive text with the subject of the image. If alt text is controlled by users, we



suggest that you recommend that your users follow the [guidelines outlined by Google](#). This includes helpful information on file names that can be useful to boost your SEO.

## 10.2 Recommendations

- Review images on the site to make sure they are not too large and/or taking up too much real estate.
- Review the reports we've provided on site speed from [Google](#) and [GTmetrix](#) and act on recommendations.
- Create useful ALT text, keeping both users and search engines in mind. For user-generated ALT text, consider providing a short video or tutorial on creating ALT text (and titles and descriptions). Putting links directly on the page to these materials increases the chance that users will view them.

## 11.0 Video

Video Issues	
Implementation Priority	3
Content Support	Yes
Technical Support	Yes

Videos offer a great opportunity to earn visibility and additional organic traffic with video or universal search results.

Below are some video best practices to keep in mind to maximize impact and reach in the search engines:

- Optimize all the video related elements — titles, description, categories, keywords, etc. — with your target keywords, not only in the site but also in popular video platforms, such as YouTube or Vimeo. Just keep in mind that putting your videos on YouTube will sabotage your own site's traffic. However, YouTube is the second largest search engine, so oftentimes it's worth losing out on the traffic to your site.
- Use [structured markup](#) to mark up your videos.
- Add videos to your XML sitemap following the [multi-format](#) sitemap specification or create a new specific [video sitemap](#) and submit it to the search engines.

If you're using YouTube to host your videos, there are a number of ranking signals YouTube uses for its videos. You can see them in the infographic below by [Tag SEO](#): [screenshot here]

## 11.1 Issues

### *YouTube channel is not optimized*

[This resource published on Search Engine Watch](#) is a great primer in how to use tags on YouTube. SampleSite.com should be using tags liberally as they're good for search as well as good for users. The YouTube channel should also list all other social media profiles.

### *Missing call to action for people to comment on the video*

YouTube makes it easy to add [annotations](#) that encourage viewers to like videos and subscribe to your channel. You can add annotations that prompt viewers to also comment on your videos. It's important to vote up and engage with people who comment on your videos. Video responses show YouTube that your video is popular and relevant. Replying demonstrates you're approachable.

### *Missing links to other YouTube videos in annotations*

These links can be added using annotations with links to your other videos. There's anecdotal evidence that YouTube favors videos that link to other YouTube assets within annotations.

### *Lack of [keyword-rich] playlists*

If you have videos that naturally cluster around a particular topic, you should consolidate these into [playlists](#). Well-architected playlists perform well in search engines. They should also be embedded on your site.

### *Missing link to site in YouTube description*

YouTube allows you to include links in your descriptions. This is a great way to nudge people to visit the blog post your video is embedded in for more information. A few tips:

- Include the link early in the description.

- Make sure you include the protocol (http or https), or the link won't be clickable.

### *Missing link on site to video channel*

It goes without saying that if your brand has a YouTube channel, you should include a link to your YouTube page in your site template. Even better: Consider adding the ability to subscribe to your channel right from your page/site using a Subscribe button.

You can configure the code for your site from the [Google Developers site](#). (Scroll down to [Configuration options](#) to set yours up. No developer resources necessary.)

### *Site isn't uploading native video to Facebook*

We've observed that SampleSite is uploading YouTube videos to its Facebook page. We recommend uploading videos directly to Facebook.

Search Engine Journal [ran a test](#) to see if Facebook was favoring native videos over videos embedded from third-party video embeds, such as those embedded from YouTube. Their study concluded that native videos received twice the reach, get twice as many likes, three times more shares, and seven times more comments.

## **11.2 Recommendations**

- Optimize the YouTube channel. Add additional social media profiles to your YouTube channel.
- Optimize each video. Most descriptions are too short. Each video description should have a link back to page on the website with like content with a link. Videos should be grouped into content specific playlists.
- Increase output of engaging video content. Create an XML video sitemap when video output increases.
- Create and upload transcripts for videos, use YouTube annotations at the end of the videos to drive traffic back to your site.
- Explore uploading native video to Facebook.

## 12.0 Titles

Title Issues	
Implementation Priority	
Content Support	Yes
Technical Support	Yes

Title tags (technically called title elements) define the title of a document and are required for all HTML/XHTML documents. It is the single-most important on-page SEO element (behind overall content) and appears in three key places: the title bar of a browser, search results pages, social sites:

As a ranking factor in organic search and a highly visible element to users in the search engine result pages (SERPs), it's important to make sure that a site's titles meet certain criteria:

- They're relevant, descriptive, and unique for each page.
- They read effortlessly. Titles with keywords separated by pipe characters, for example, aren't good for social sharing or click throughs from search results.
- They're different from the meta description and offer complementary information.
- They include the most important keywords for each specific page, preferably toward the beginning of the title (but aren't keyword stuffed!).
- They include the name of the site at the end of the title to strengthen its branding.
- They're concise since search engines only show the first ~70 characters.

We acknowledge that the majority of page titles on the SampleSite.com site is user-generated and therefore quality control is challenging. We recommend addressing the issues that on pages that are directly controlled by SampleSite (especially top landing pages and money pages) and strategize how to avoid issues with user-generated content in the future.

You can find a report of all the title tag issues in the [page-titles-audit-data workbook](#).

## 12.1 Issues

### *Pages with missing titles*

Some pages on SampleSite are missing title tags. We've included a report in the [audit-data spreadsheet](#) for your review. If you don't already, we recommend that you make the title field mandatory for campaign pages.

### *Pages with long titles*

Best practice is to keep page titles under 70 characters (or 500 pixels) to be safe. Any larger than that and page titles could be truncated in the search engine result pages (SERPs). There is a correlation between truncated titles and lower click through rates (CTR).

If you don't already, consider establishing a maximum character limit (and perhaps also a minimum) for user-generated page titles.

This screenshot shows titles that are too long that get truncated in Google search results (as well as the presence of duplicate content): [screenshot here]

### *Pages with short titles*

SampleSite.com has 551,798 with page titles that are shorter than 30 characters. Excessively short page titles can fall short of adequately communicating what content a viewer can expect to see when s/he clicks through.

### *Duplicate page titles*

Pages with duplicate titles are oftentimes good indicators of duplicate content (which was covered in the [Canonicalization section of the audit](#)).

### *Odd capitalizations in titles*

Page titles typically use title case. Capitalizing words just to emphasize them can make your page titles distracting and give your site an amateur appearance. Also avoid using special characters such as those used by Microsoft Word. Example: Sample Site &#x2013; An Online Website | SampleSite - SampleSite. (Unrelated issue: replace the text in red with the appropriate HTML character entity).

## **12.2 Recommendations**

- Again, we understand that controlling much of the title tag content is difficult for SampleSite because of user-generated content. As much as possible, use best practices with page titles and help your users do so too. Consider creating a video or tutorial to walk users through best practices of setting up a campaign page and linking to it from the setup page. Also consider applying a maximum and minimum title length. Make titles mandatory if you aren't already doing so.

## 13.0 Meta Descriptions

Meta Description Issues	
Implementation Priority	3
Content Support	Yes
Technical Support	Yes

Although meta descriptions are not a ranking factor, they're highly visible to users in the SERPs and can affect click through rates. Google highlights not only search query terms but also synonyms in meta descriptions, making these keywords pop off the page for searchers.

With this in mind, you should keep a few things in mind when writing meta descriptions:

- They should have a compelling call to action to lure the searcher to click through.
- They should be unique and describe the content of the page well.
- They should be no more than 156 characters long since search engines truncate descriptions beyond this point (and sometimes before this point since Google uses pixel width instead of characters now).

Just as with page titles, we understand that user-generated meta descriptions present a challenge for SampleSite. We recommend applying the same strategy for meta descriptions as with title tags.

You can find a report of all the meta description issues in the [meta-description-audit-data workbook](#).



## 13.1 Issues

### *Pages with missing meta descriptions*

Any page that is intended to be public facing should have a meta description. There are 854,131 pages with missing meta descriptions on SampleSite.com.

Some of the key pages on SampleSite are missing meta descriptions. Example:

<https://www.SampleSite.com/orders/new> and <https://www.SampleSite.com/us/story/new>.

When considering adding meta descriptions, these pages and other top pages directly controlled by SampleSite should be a priority.

### *Pages with long meta descriptions*

The long meta descriptions tab of the [meta-description-audit-data](#) workbook lists pages that have long meta descriptions. Long descriptions will be truncated. This reduces their readability and click through rates.

### *Pages with short meta descriptions*

The short meta descriptions tab lists pages that have short meta descriptions. Care should be taken to write descriptions that are adequately descriptive and compelling.

### *Pages with duplicate meta descriptions*

Not only are there duplicate meta descriptions across the site, there are also meta descriptions duplicated on the same page. Under no circumstances should there be instances of multiple meta descriptions on a page.

We notice also that there are duplicate meta descriptions across pages on different subdomains. This is probably related to the fact that there is duplicate content across subdomains. We discuss this in [previous section](#) in this report. As much as possible, meta descriptions should be unique across the site.

## 13.2 Recommendations

- As we pointed out with title tags, we understand that controlling user-generated content is a challenge. See the recommended strategy we suggested with [title tags](#).
- Assess and remove duplicate meta descriptions that occur on the same page.
- Ensure that meta descriptions that you control have a compelling call to action.

## 14.0 Site Speed

Site Speed Issues	
Implementation Priority	1
Content Support	No
Technical Support	Yes

On April 9<sup>th</sup>, 2010, [page load speed officially became a part of the Google search ranking algorithm](#). For usability reasons, best practices dictate that a web page should load within 1-2 seconds on a typical connection. However, according to Search Console data, a load time of 1.4 seconds is the threshold between a fast page and a slow page. That means, ideally, that every page on your website should load in 1.4 seconds or less, to receive the maximum SEO benefit for fast-loading pages. As sites get faster, user expectations change. So user expectation would even go lower.

Besides the ranking benefits, there are crawling benefits for faster sites. As discussed in the [Crawl section of the audit](#), each site is assigned a crawl budget. When search spiders have exhausted that budget, they move on. There is a strong correlation between improving page speed and the search engines' ability to crawl more pages.

Google gathers page load time data through actual user experience data collected with the Google Toolbar, Chrome, and Android and may also be combining that data with data collected as Google crawls a website. As such, page load speed in terms of the ranking algorithm is being measured using the total load time for a page, exactly as a user would experience it.

You can now view samples of page load times right in Google Analytics. It's especially helpful to segment those page load times out by country as you can see in [this report](#).

The overall page speed for the past 3 months hovered in the neighborhood of 10 seconds. This is considered a very high load time. [screenshot here]

## 14.1 Issues

### *Page load times of top landing pages high*

The average page speed for SampleSite for the past 3 months for the top landing pages is 6.66 seconds. Here is a link to [the report](#).

This is an average for the top landing pages only. Page speed for the site as a whole is slower.

We created a [video](#) to give you an idea of how the page speed on SampleSite.com stacks up to two of its competitors, Indiegogo and Kickstarter. Kickstarter has the fastest page speed at 4.4 s, followed by Indiegogo at 8.6 s. SampleSite tops out at 12.1 s.

### *Page load times of money pages lag in comparison to competitors*

As we did with the homepage, we created a [video](#) to illustrate how SampleSite compared with Kickstarter and Indiegogo in terms of page speed. This time we tested the start a fundraiser page which is a key a money page for the site. This time, Indiegogo had the lowest page speed at 4.0 s, followed by Kickstarter at 5.3 s. Once again was last with 6.1 s.

### *Page Caching*

The SampleSite.com site does not leverage page caching. What caching does is reduce the page load time of your pages by storing commonly used files from your website on your visitor's browser. This avoids potentially time-consuming queries to your database. It also helps lower server load by up to 80%.

Caching most visibly benefits high-traffic pages that access a database but whose content does not change on every pageview. Common caching methods include [Alternative PHP Cache](#), [Quickcache](#), and [jpcache](#). Caching mechanisms also typically compress HTML, further reducing page size and load time. Get a full list of resources you can cache from the GTmetrix report for the SampleSite.com site. You can read more about the benefits of caching [here](#).

### *Blocking scripts*

SampleSite.com has 2 blocking scripts and 2 blocking CSS resources. This can cause a delay in page load and rendering along with the issues mentioned above.

### *Site not using gzip compression*

The SampleSite.com site does not appear to be using gzip compression. You should compress your HTML to reduce your page size and page loading times. This will help your site retain visitors and increase pageviews. Consider using GZIP to compress your HTML. The GTmetrix report for the SampleSite.com site contains specific resources it recommends for gzip compression that will reduce the size of the files listed by 74%. [screenshot here]

### *Site uses inline CSS*

Inline styles are CSS styles that are specified within the page code itself, as opposed to inside an external style sheet. Repeated use of inline styles leads to unnecessary code bloat and slower page loading. Using an external CSS style sheet that references these styles will lead to smaller code and faster page loading.

An inline CSS property is added by using the style attribute for a specific tag. By mixing content with presentation you might lose some advantages of the style sheets. It is good practice to move all the inline CSS rules into an external file in order to make your page "lighter" in weight and decrease the code-to-text ratio.

### *Site uses inline JavaScript*

The SampleSite.com homepage uses 6 inline Javascript properties. If your site is using the same JavaScript code across multiple pages, you should consider using an external JavaScript file so that it will be cached by the user's browser for each of those page views. This will reduce your page's HTML size, as well as make your code easier to maintain.

### *More recommendations from GTmetrix and PageSpeed Insights reports*

You can view more recommendations to reduce page load time for the SampleSite.com site by navigating to these reports: [GTmetrix report](#) and [Google PageSpeed Insights report](#).

## **14.2 Recommendations**

- Improve page speed by following the recommendations outlined above as well as in the [GTmetrix report](#) and [Google PageSpeed Insights report](#).
- Ensure that content takes less than 4 seconds to render (when using prerender.io) or that content that will not get indexed by Google.

## 15.0 Mobile

Mobile Issues	
Implementation Priority	3
Content Support	No
Technical Support	Yes

Mobile is the future of the Internet, and the future is here. In May 2015 [mobile finally eclipsed desktop traffic](#). Google has doubled down on trying to ensure a positive user experience for users it sends to mobile sites and apps.

To that end, Google released its mobile-friendly algorithm update on April 21, 2015, and sent many site owners and marketers scrambling to get their sites—or at least key landing pages — mobile ready. Google released some helpful resources to help site owners and marketers create mobile-friendly sites, like the [Mobile-Friendly Websites guide](#).

There are a few key facts marketers and site owners should know about this update:

- It is assessed on a page-by-page basis in real time.
- Google’s mobile-friendly assessment determines if a site is mobile or it’s not. There’s no in between.
- Google released a [Mobile Friendly Tool](#) to help marketers assess if their key landing pages were mobile friendly.
- If a page passes the mobile-friendly test, it could be rewarded with a snippet in mobile search results. [screenshot here]

### *Mobile visitors bounce at a higher rate*

The bounce rate (the percentage of sessions in which the person left your site from the entrance page without interacting with the page) for mobile visitors is higher than that of

desktop and tablet users. While a high bounce rate is not always negative, it can indicate that users are not satisfied with the page speed, the content, or other factors. Further investigation should be done to determine the cause.

## 15.2 Recommendations

- Ensure that the mobile configuration follows best practices, and all mobile URLs are 301 redirecting back to the main responsive site.
- Re-review bounce rate of mobile visitors when site speed, content, and other enhancements are made.



## 16.0 Social

Social Issues	
Implementation Priority	3
Content Support	Yes
Technical Support	Yes

Social media plays an essential role in 21<sup>st</sup>-century society. Users enjoy the ability to interact with friends and family, get involved in a plethora of activities that fit their interests, receive support, find a potential partner, follow news and current events, interact with celebrities, review businesses, etc.

Organizations can benefit from this craze by being astute, approachable, sensitive, witty, and helpful ... to name just a few. Conversely, organizations can (and have!) do damage (sometime irreparable damage) to their brands wielding their social media swords irresponsibly. But most frequently sites allow minor technical issues and workflow inefficiencies to choke out the effectiveness of their social media efforts. This section of the audit will address some of these issues.

### 16.1 Issues

SampleSite.com uses social media such as Facebook and Twitter on a regular basis. Facebook is the top social medium that people interact with.

Other social media assets include YouTube (semi-active), Instagram (not active since 2014), Google+ (not active since 2014).

We also note that SampleSite allows users to post Instagram feed on fundraiser campaign pages. This is a nice touch and can help with user engagement and social sharing.

We'd like to see SampleSite increasing their activity on social media channels as well as crafting engaging, compelling content on a regular basis. Using a familiar, conversant tone helps to attract users and keep coming back which can increase social shares and mentions. Other

things SampleSite could do is track their brand mentions and acknowledge them with positive feedback. Also, SampleSite could start using their [SampleSitesupport Twitter account](#) and linking to it from other social channels. This account has not been used since 2015. It can show users that SampleSite is there to help should users need it.

### *Site doesn't take advantage of Twitter Cards markup*

Twitter Cards make shared content on Twitter more attractive and potentially more engaging. Here is a tweet from SampleSite (no Twitter card): [screenshot here]

And this is what a tweet looks like with a [player card](#):

A few notes:

- Make sure the image in your post that the Twitter card will pull from is at least 120 x 120px (rectangular images work best) and under 1mb in size.
- Buffer published an [excellent Twitter Cards guide](#) you should check out. Also see [Twitter's own page on Twitter cards](#). [screenshot here]

### *Facebook Open Graph markup*

The news.SampleSite.com blog site uses Open Graph markup. Some other pages on the site do as well. Consider using on SampleSite.com money pages such as <https://www.SampleSite.com/gift/orders/new>). See <http://ogp.me/> for ideas about additional opportunities to use Open Graph markup to increase social engagement.

One other thing to note is that Google+, LinkedIn, and Pinterest also use Open Graph markup, so setting this up is like scoring a buy-one-get-three-free sale.

### *Use Native Social media follow and "like" buttons*

The site is currently using ShareThis social sharing buttons.

We would recommend using the native social media buttons (the Facebook "like", Twitter "tweet" and Google+ "+1" buttons) on each and every page on the site. Additionally, the site

should use the native channel/subscribe/follow buttons on their site to drive visitors to their social channel.

Those buttons (provided by the social platforms) allows you to encourage website visitors to “like”/follow your Facebook page (for example) without leaving your website AND personalizes that user’s experience when they next visit Facebook by increasing the possibility that they will see your brand posts in the Facebook feed. This personalization happens on all of the social media channels, and we would recommend adding the YouTube subscribe button, the Google+, Facebook, and Twitter follow buttons.

Here’s an example of the native social media follow buttons: [screenshot here]

### *Content that’s performed best in social*

Below are the pages from the site that have performed best on social media over the last year. One is the personal page on go.SampleSite.com. [screenshot here]

Look carefully at the social content that has the most engagement. This Facebook post is a good example of content that performed well:

### *Social networks that perform best on your site*

Below are the social networks that drive the most inbound traffic over the past month: [screenshot here]

Most of this traffic is to individual pages, and when you look at just the homepage it seems as though SampleSite might be too reliant on its Facebook strategy for referrals and should diversify its social media promotion:

### *Optimize social media profiles*

SampleSite should merge its two Google + profiles. This one is a local profile <https://plus.google.com/+SampleSite> And this is the one linked to the YouTube channel: <https://plus.google.com/106099094532>. Promotional posts should be on the brand channel.

Instructions for how to merge the two but still keep the Google map data is here:

<http://localu.org/blog/google-now-allows-brand-pages-become-local-pages/>

SampleSite should then optimize their new merged Google+ profile by adding links to all of the other SampleSite profiles, as this is your account with Google. Additionally, the site should also be linked to the Google+ profile using rel=publisher. As Google says in their document:

*Linking your brand page and website helps you connect with your friends, fans, and customers. It also provides Google with information that helps determine the relevancy of your site to a user query on Google Search.*

Instructions for how to link the site to its Google + profile are here:

<https://support.google.com/business/answer/4569085?hl=en>

Additionally, each Google+ post should include descriptive text and appropriate hashtags.

The Pinterest profile should be verified. Instructions are here:

<https://help.pinterest.com/en/articles/confirm-your-website>

## 16.2 Recommendations

- Use the native social media follow and like buttons.
- Link your Google+ page to your website using rel=publisher.
- Optimize your YouTube channel.
- Diversify social channels. Don't rely just on Facebook and Twitter.
- Identify what content does well on social channels and what doesn't. Learn how to create more engaging content from this.
- Make sure that the site uses Twitter Cards and Open Graph markup across the site.