

THE BUSINESS IMPACT OF WEBSITE SCRAPING: IT'S PROBABLY BIGGER THAN YOU THINK — HERE'S WHY

May 2020

Derek E. Brink, CISSP

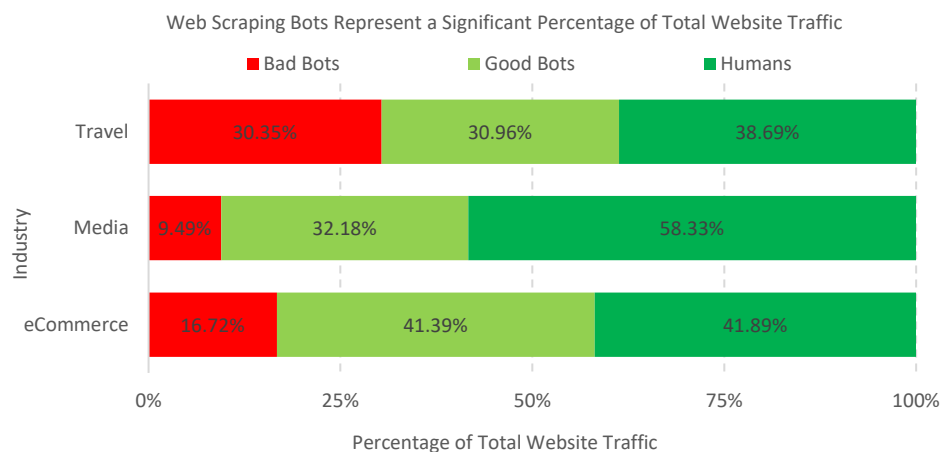
Vice President and Research Fellow, Information Security and IT GRC

From the perspective of website owners, **website scraping** is a gain for your competitors — at your expense. Website scraping **bots** represent a surprisingly high percentage of total website traffic, which leads to both **higher costs** (for *website infrastructure*, and *website marketing*) and **lower revenue** (from current and future buyers). Aberdeen’s analysis quantifies the business impact of website scraping, for three selected industry sectors — **eCommerce**, **Travel**, and **Media** — and puts it in perspective in comparison to overall industry profitability.

Website Scraping: Friend, or Foe? Answer: Yes

Bots represent a significant percentage of total website traffic. Based on its visibility into website traffic for hundreds of sites, a leading provider of application protection solutions gives us fact-based insights into just how much: Roughly 40% to 60% of total website traffic — with differences based on industry — is from actual human users (see Figure 1).

Figure 1: Just 40% to 60% of Website Traffic is Actual Human Users



Source: Analysis based on empirical data from PerimeterX; Aberdeen, May 2020


The rest is comprised not of humans, but of bots.

► Bots

A **bot** refers to software which is commonly harnessed in conjunction with other bots on hundreds or thousands of host systems (a **botnet**), to work together towards some common objective.

► Website Scraping

Website (or web) scraping refers to the process of extracting data from a public-facing website — such as *content*, *pricing*, *product information*, *inventory levels*, and *customer reviews* — typically at speed and scale, i.e., automated by using a bot or botnet.



From the perspective of the website owner, not all bots are malicious. Where would we be, for example, without the “good” bots used by Internet search engines to find and index our website content? These bots make it easier and more convenient for human users — our current and future customers — to find the products and services they’re interested in.

The bots with malicious intent are the ones we should be concerned about. As we have seen, empirical data shows that **bad bots represent roughly 10% to 30% of total website traffic**, depending on the industry. The important question is, have we properly understood and quantified the potential *business impact* that bad bots represent?

If we didn’t know otherwise, a quick search on **web scraping** might lead us to worry more about missing out on its purported *benefits*, as opposed to protecting against its downside *risks*. Consider these illustrative titles that result from such a search:

- ▶ *The Future of Web Scraping, and How It Matters*
- ▶ *5 Reasons Why Web Scraping May Benefit Your Business*
- ▶ *Web Scraping Tips and Tricks for Efficient Business*
- ▶ *A Guide to Web Scraping Without Getting Blocked*
- ▶ *Best Web Scraping Tools: Top 15 Web Scrapers in 2020*

Again, let’s look at this from the perspective of the website owner. Notice that the first three examples hint at the *benefits* of being the one doing the scraping. On the other hand, the fourth example acknowledges that website owners generally don’t want their public-facing data to be scraped by competitors, and may take proactive technical steps to *block* it. They may also look to the law for protection, although high-profile legal challenges thus far have ruled that scraping public-facing information is not illegal.

The fifth example highlights the fact that website scraping is highly popular, and supported by a number of automated tools that are available to make the process faster and easier.

Website Scraping is a Zero-Sum Game

When it comes to scraping my website data, *your gain is my loss*. As a hypothetical example: Suppose one furniture and home-goods retailer

Website scraping is a **zero-sum game**. When it comes to scraping my website data, *your gain is my loss*.

automates the scraping of the websites of a large competitor, to help optimize its own mix of products, providers, and pricing. By doing so, this contributes to generating more revenue for itself. In this scenario, one competitor’s gain is another competitor’s loss.

High-level factors of the business impact of website scraping are summarized in Table 1, from both perspectives.

Table 1: Website Scraping is a Zero-Sum Game

| Your Competitors Gain | Your Loss |
|---|---|
| <p>Reduced costs</p> <ul style="list-style-type: none"> Gain operational efficiencies from automation (e.g., bots) — higher scale, at lower cost | <p>Increased costs</p> <ul style="list-style-type: none"> Wasted expenditure on website marketing Wasted expenditure on website infrastructure |
| <p>Increased revenue</p> <ul style="list-style-type: none"> Expand / optimize your product mix, content Compare / optimize your pricing Increase traffic from target buyers Gain new / repeat buyers | <p>Decreased revenue</p> <ul style="list-style-type: none"> Loss of competitive advantage (product mix, content, pricing) Decreased traffic from target buyers Loss of existing / new / repeat buyers |

Source: Aberdeen, May 2020

To better understand how website scraping translates to *wasted expenditure on website marketing*, for example, consider that website marketing expenditures typically include a combination of **Search Engine Optimization (SEO)**, **Search Engine Marketing (SEM)**, and **Web Analytics**:

- ▶ *SEO* refers to strategies to increase your website traffic by raising its visibility in Internet search engine results pages, primarily through organic techniques. For example: By stealing well-written, keyword-optimized product descriptions, content-scraping bots can help your competitors leverage your hard-earned investments to improve their own SEO rankings — while potentially penalizing yours, and making it harder for your customers to find you.
- ▶ *SEM* refers to strategies to increase your website traffic by raising its visibility in Internet search engine results pages or other website pages, primarily through various forms of paid advertising. Paying for digital ads on pages being visited and scraped by bots

— as opposed to your potential buyers — is also a waste of your marketing investments.

- ▶ *Web Analytics* refers to capabilities for measuring, analyzing, and gaining insights into your website traffic, to help make better-informed decisions about how to optimize your website experience for legitimate buyers. Bad bot traffic skews the results of Web Analytics, which squanders your investments in doing these activities in the first place — and in the worst case, can easily lead you to make suboptimal decisions, faster and with even greater confidence.

In all three of these areas, the common theme is that website marketing represents a significant investment — to the tune of between 5% to 20% of top line website revenue, depending on the industry.

These investments are designed to generate website traffic, and ultimately to lead to new / repeat buyers — and no one wants to pay for traffic from bad bots, or incur the negative business impact they create.

The Business Impact of Website Scraping: It's Probably Bigger Than You Think — Here's Why

To quantify the business impact of web scraping bots, Aberdeen developed a simple **Monte Carlo** analysis based on the best available estimates for the *range (lower bound, upper bound)* and *shape (distribution)* for each of just six key factors:

- ▶ Industry sector (e.g., eCommerce, Media, Travel)
- ▶ Website contribution to annual revenue (e.g., \$100M / year)
- ▶ Percentage of website traffic attributable to web scraping bots (based on empirical data from application protection solution provider PerimeterX)
- ▶ Percentage of annual revenue spent on website infrastructure
- ▶ Percentage of annual revenue spent on website marketing
- ▶ Percentage of annual revenue lost as a result of website scraping (e.g., pricing, content)

Monte Carlo models are a proven, widely-used approach to quantitative analysis, and can help to provide useful insights to help make better-informed business decisions in the face of inherent uncertainties.

▶ Sidebar: An Example of Skewed Web Analytics

In the Travel industry, the number of visitors to a website divided by the number of people who actually make a reservation is commonly referred to as the **Look-to-Book ratio**.

Traditionally, travel-related websites have used Look-to-Book as one way to gauge the return on investment of their website marketing expenditures, which are designed to turn casual browsers into confirmed bookers.

In this hyper-competitive segment driven by price and schedule, visits by price-scraping bots are driving Look-to-Book ratios sky high — e.g., for airline sites, the global average abandonment rate is currently about 88% (*Source: SaleCycle, 2020*).

Computations can be carried out for many scenarios (e.g., ten thousand), each of which uses a random value from the estimated ranges and shapes for the inputs — as opposed to a single calculation based on a single, static value such as “the average cost of a data breach is \$148 per record.”

The results of these computations are likewise not a single, static value — but are in the form of a *range of possible values*, along with the *associated likelihoods*. From this, we can readily describe both *how likely* and *how much business impact*, i.e., the **risk** of website scraping, as risk is properly defined.

Finally, one way to put the business impact of website scraping in perspective is to consider it in relation to the overall website profitability for each given industry sector. For the purposes of this analysis, Aberdeen has used the range of annualized **EBITDA** (based on the trailing twelve-month period, or TTM) over the last five quarters as the yardstick for overall website profitability against which the total business impact of website scraping should be compared.

With that context, Aberdeen’s quantitative analysis of the business impact of website scraping is summarized in Figure 2 (eCommerce), Figure 3 (Media), and Figure 4 (Travel).

eCommerce

For the **eCommerce** sector, Aberdeen’s analysis estimates that the annual business impact of website scraping is **between 3.0% and 14.7%** (**median: 8.1%**) of annual website revenue.

In comparison, over the past five quarters the overall profitability (TTM EBITDA) for the eCommerce sector has ranged **between 10.1% and 12.0%** of annual website revenue.

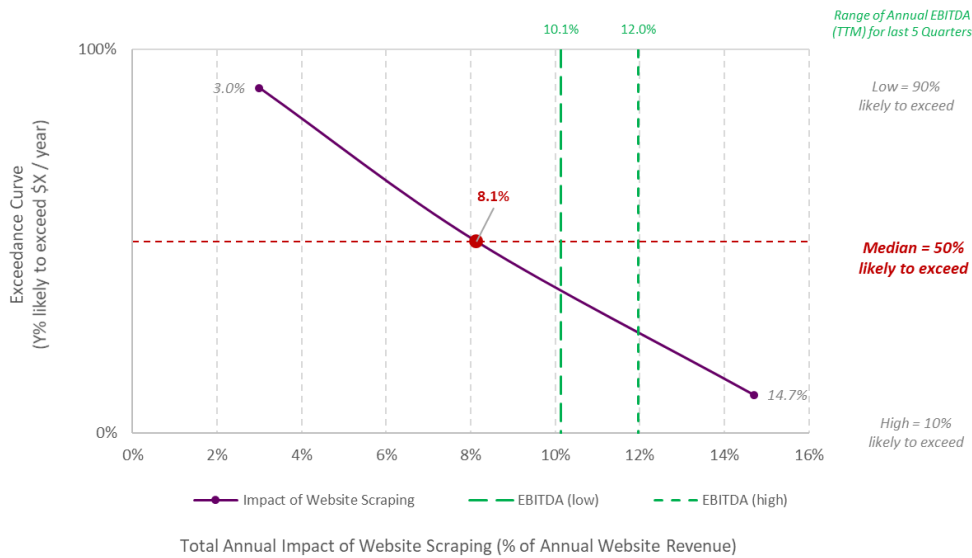
This provides invaluable insights for website owners in eCommerce about the *risk* of website scraping:

- ▶ The **median** annual business impact of website scraping is **as much as 80%** of overall website profitability
- ▶ **For every \$1** of website profitability, there’s an **80% likelihood** that the annual business impact of website scraping will be **between a range of \$1.23 to \$1.46**

Is this a level of risk they are willing to accept?

Earnings Before Interest, Taxes, Depreciation, and Amortization (EBITDA) — commonly expressed as a percentage of top-line website revenue — is a commonly used measure of overall profitability.

Figure 2: Annual Impact of Website Scraping (eCommerce)



Source: Monte Carlo analysis; Aberdeen, May 2020

Media

For the **Media** sector, Aberdeen’s analysis estimates that the annual business impact of website scraping is **between 3.0% and 14.8% (median: 7.9%)** of annual website revenue.

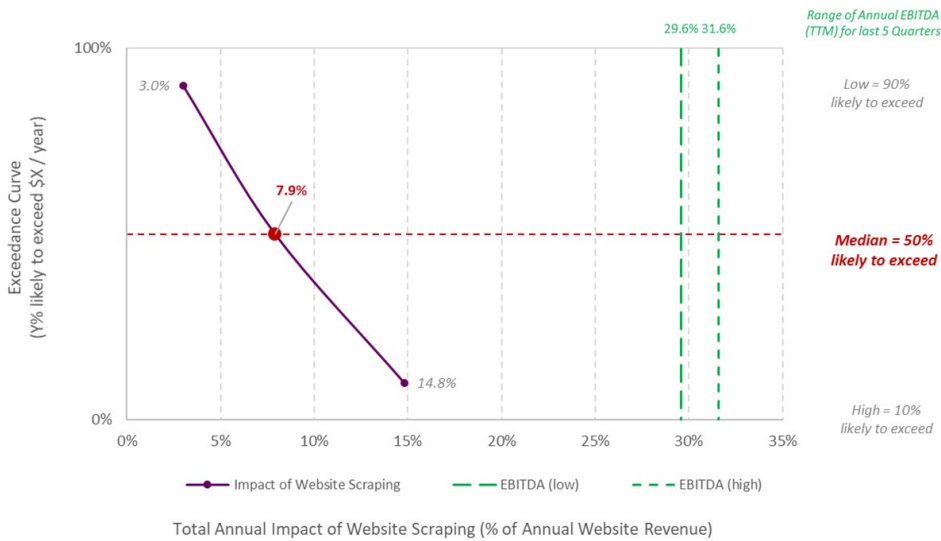
In comparison, over the past five quarters the overall profitability (TTM EBITDA) for the Media sector has ranged **between 29.6% and 31.6%** of annual website revenue.

This provides invaluable insights for website owners in Media about the *risk* of website scraping:

- ▶ The median annual business impact of website scraping is **as much as 27%** of overall website profitability
- ▶ **For every \$1** of website profitability, there’s an **80% likelihood** that the annual business impact of website scraping will be **between a range of \$0.47 to \$0.50**

Is this a level of risk they are willing to accept?

Figure 3: Annual Business Impact of Website Scraping (Media)



Source: Monte Carlo analysis; Aberdeen, May 2020

Travel

For the **Travel** sector, Aberdeen’s analysis estimates that the annual business impact of website scraping is **between 6.0% and 18.3%** (**median: 11.6%**) of annual website revenue.

In comparison, over the past five quarters the overall profitability (TTM EBITDA) for the Travel sector has ranged **between 14.9% and 20.5%** of annual website revenue.

This provides invaluable insights for website owners in Travel about the *risk* of website scraping:

- ▶ The **median** annual business impact of website scraping is **as much as 78%** of overall website profitability
- ▶ **For every \$1** of website profitability, there’s an **80% likelihood** that the annual business impact of website scraping will be **between a range of \$0.89 to \$1.23**

Is this a level of risk they are willing to accept?

Figure 4: Annual Business Impact of Website Scraping (Travel)



Source: Monte Carlo analysis; Aberdeen, May 2020

Website Owners: You Need to Make a Better-Informed Business Decision About Your Risk from Website Scraping

Aberdeen’s quantitative analysis helps to bridge the gap from *technical descriptions* of website scraping, to also include better-informed *business conversations* about its corresponding business impact — i.e., not only “what could happen,” but also “why it matters.”

By your own organization’s technical staff doing this kind of analysis for the risk of website scraping, it will be fulfilling both its traditional role of *subject-matter experts* (the “what”), as well as its emerging and critical role of *trusted advisors* (the “so what”).

Quantitative models such as these are not intended to be *perfect* — they are intended to be *useful*, for helping website owners to make better-informed business decisions about the risks of website scraping.

Ultimately, the business decision regarding what to do about the risk of website scraping — i.e., *avoid*, *accept*, *transfer* to another party, or take steps to *manage* to an acceptable level — is up to the website owners, not the technical staff. Given the exact same context, one organization may decide to accept the risk, while another may decide to take steps to reduce it.

In Aberdeen’s view, the business impact of website scraping is probably bigger than you think, and justifies an investment in countermeasures.

Related Research

- ▶ *Open Source Software: To Get More Value, Manage Your Risks;* January 2020
- ▶ *Mobile Device Security: No Perimeter? No Problem;* November 2019
- ▶ *The Value of Cloud Data Protection for Remote Office / Branch Office;* July 2019

About Aberdeen

Since 1988, Aberdeen has published research that helps businesses worldwide to improve their performance. Our analysts derive fact-based, vendor-neutral insights from a proprietary analytical framework, which identifies Best-in-Class organizations from primary research conducted with industry practitioners. The resulting research content is used by hundreds of thousands of business professionals to drive smarter decision-making and improve business strategies. Aberdeen is headquartered in Waltham, Massachusetts, USA.

This document is the result of primary research performed by Aberdeen and represents the best analysis available at the time of publication. Unless otherwise noted, the entire contents of this publication are copyrighted by Aberdeen and may not be reproduced, distributed, archived, or transmitted in any form or by any means without prior written consent by Aberdeen.