

Maximizing the Value of Web Data

From Extraction to Transformation to Insights

Executive Summary

Tapping into web data represents a competitive imperative for organizations across a wide range of industries, including retail, manufacturing, equity research, insurance, data science, and marketing.

However, to realize the potential of web data, you need a practical way to extract the data and make it actionable. Import.io provides a platform that automates web data extraction and enables you to transform the data into exactly what you need for visualizing and analyzing data trends and characteristics.

By using web data in critical decision processes, organizations gain the intelligence needed to optimize:

- Product and pricing
- Investment decisions
- Risk management
- Mergers and acquisitions
- Machine learning and artificial intelligence models
- and much more...

Maximizing the Value of Web Data

INTRODUCTION: THE MASSIVE SCALE AND POTENTIAL OF WEB DATA

With billions of pages of data, the web currently represents an immensely valuable source of intelligence for informing strategic business decisions.

Following are just a few of the ways that web data contributes to the evolution of industries:

Retail:

- Monitor competitor and marketplace prices to enhance pricing strategies and more consistently beat the competition.
- Mass download product images and descriptions from manufacturing websites to populate the latest information on an online marketplace.
- Monitor customer sentiment about products, services, and brands.
- Determine inventory completeness and opportunities based on competitor offerings, find gaps and fill them.

Manufacturing:

- Automate the extraction of web data to streamline the identification of the lowest priced suppliers, and stay on top of pricing changes.
- Track how retailers are promoting their goods, and ensure they are complying with established policies, such as minimum advertised price requirements.

Equity research:

- Improve investment models by leveraging data from such non-traditional sources as industry blogs, social media sites, news aggregation sites and more.
- Gather data from a variety of market information sources for most current intelligence.
- Extract data from publicly available financial statements to streamline research.

Insurance:

- Track reviews of products to understand quality and determine cost of insurance.
- Gather alternative data to determine insurance risk.

“Web scraping is ‘the original growth hack’ used by Fortune 500 companies to stay competitive on price, inform strategy and measure customer sentiment.”

*Source: Once Only for Huge Companies, ‘Web Scraping’ Is Now an Online Arms Race No Internet Marketer Can Avoid
Entrepreneur Magazine*

Maximizing the Value of Web Data

Data science:

Harness the mammoth volumes of web data to gather text and images and train machine learning models, such as:

- Self-driving cars.
- Disease diagnostics.
- Determining loan risk.
- Evaluating risk in hiring or partnering.
- Customer support automation.

Sales and marketing:

- Optimize efforts for defining target customer profiles, generating leads, creating data-driven content, evaluating SEO, and much more.
- Mine web data for developing differentiated content marketing assets.
- Develop high-quality prospect lists for lead generation efforts.

THE CHALLENGES OF UTILIZING WEB DATA

While the benefits of harnessing web data are obvious, determining how to start has not been so clear cut. When companies try to use internal developers, scripts, or rudimentary tool sets to scrape websites, they quickly realize this effort isn't so easy. Web data is constantly evolving, not only in terms of its content but the structure and format of websites. Once data is extracted, it often needs a lot of manual effort to standardize and parse it, so it can be fed into existing tools and processes and ultimately be useful. Using a manual process or developing one-off data scrapers may be OK for a one-time analysis or proof-of-concept, but it's not sustainable beyond that.

To fully capitalize on the promise of web data extraction, organizations need an effective platform, one that can automate not only the extraction of data from the web, but also the filtering, cleaning, and standardization of data, and the integration of this data into existing tools chains and workflows.

“Developing software is expensive. Solving problems that have been solved before, and on top of that having to build additional features, maintain the platform, and fix bugs on top of that, is very often not worth it.”

Source: *Build vs. Buy- Considerations for New Products*
Thoughtbot.com

Maximizing the Value of Web Data

THE SOLUTION: IMPORT.IO

Import.io enables you to quickly, easily, and cost effectively harness the power of web data. The company delivers a SaaS solution that can convert massive volumes of data from websites into structured data—without requiring your team to do any coding.

Import.io automates web data extraction, eliminating time-consuming, error prone manual processes and gets a steady stream of high quality web data to integrate into your processes. From hundreds of rows of data to millions, Import.io extracts the data you need. Import.io Transform enables you to clean and prepare the data for visualization and analysis using Import.io Insights. The transformed data is saved in Import.io's cloud and is kept up to date every time data is extracted.

The following sections offer more information on each of these capabilities.



EXTRACT



TRANSFORM



INSIGHTS

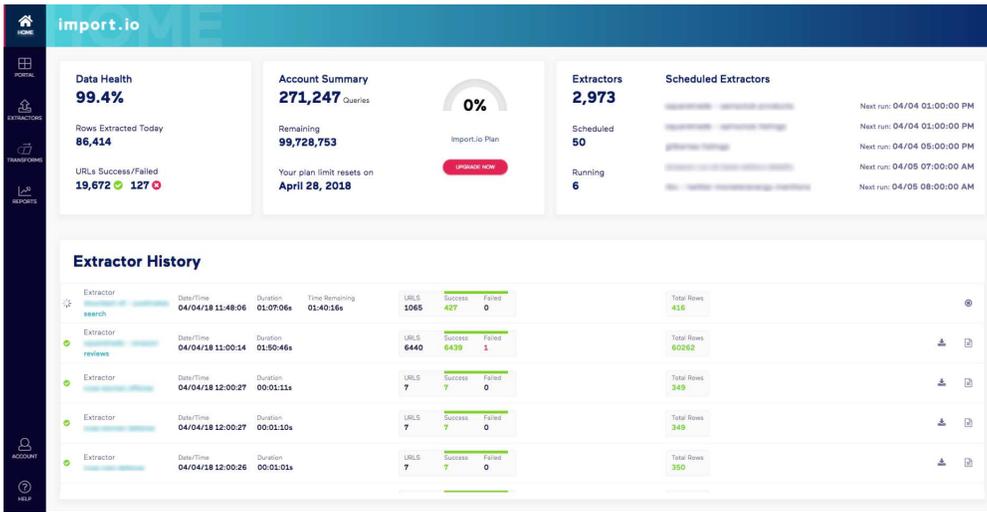
Web Data Extraction

Import.io Extract delivers capabilities that streamline web data extraction. You just enter the web page URL(s) where the desired data is located and Import.io follows the link(s) to get the page(s). If getting to your data requires a login or filling out a form, you can interact with the website and gain access to the page in Import.io. Once on the target web page, Import.io uses machine learning to determine what data you need or you can simply point and click on the data and Import.io structures it into columns.

By leveraging automated web data extraction, you eliminate time-consuming, error-prone manual processes and get high quality data. From hundreds of rows of data to millions, Import.io extracts exactly what your organization needs.

Once you have the structured data, you can download it or use Import.io APIs to provide access to a steady stream of high quality data that feeds business processes, applications, analysis tools, and visualization software.

Maximizing the Value of Web Data



Import.io Extract dashboard

Before you download the data or integrate it into your systems, you can use Import.io Transform to prepare the data.

Web Data Transformation

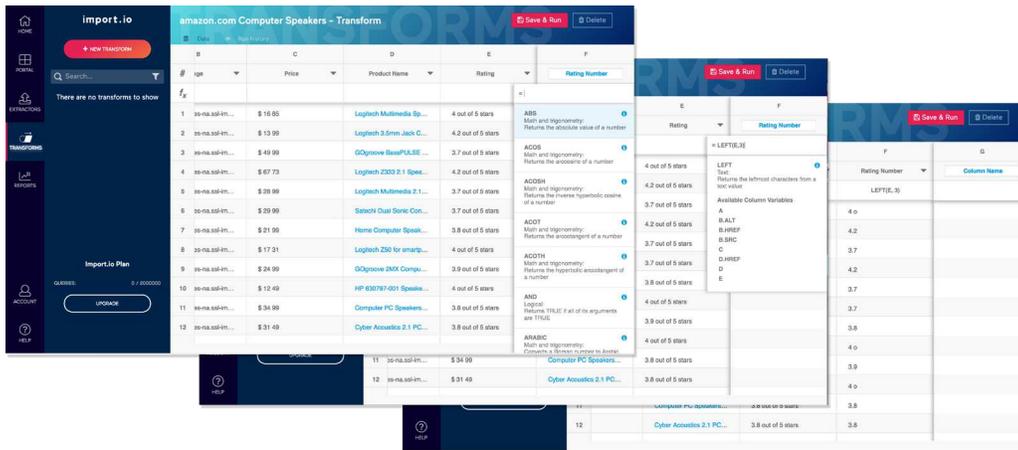
Import.io Extract delivers capabilities that streamline web data extraction. You just enter the web page URL(s) where the desired data is located and Import.io follows the link(s) to get the page(s). If getting to your data requires a login or filling out a form, you can interact with the website and gain access to the page in Import.io. Once on the target web page, Import.io uses machine learning to determine what data you need or you can simply point and click on the data and Import.io structures it into columns.

While the extraction and structuring of data from websites represent important initial steps, it is through data transformation that you can ensure you're getting the data in the form you need. Import.io Transform provides over 100 functions and formulas that are similar to those available in spreadsheet programs.

With Transform, you can clean, prepare, and normalize the web data you have extracted. For instance, removing extra spaces and symbols in a price column for consistency. You only have to set up the transformation logic once, and Import.io keeps the data up-to-date. If the data changes on a website, the next time you extract it, your transformed data is also updated to reflect these changes. This simplifies and automates data manipulation, and leaves you with clean, ready-to-use data.

Maximizing the Value of Web Data

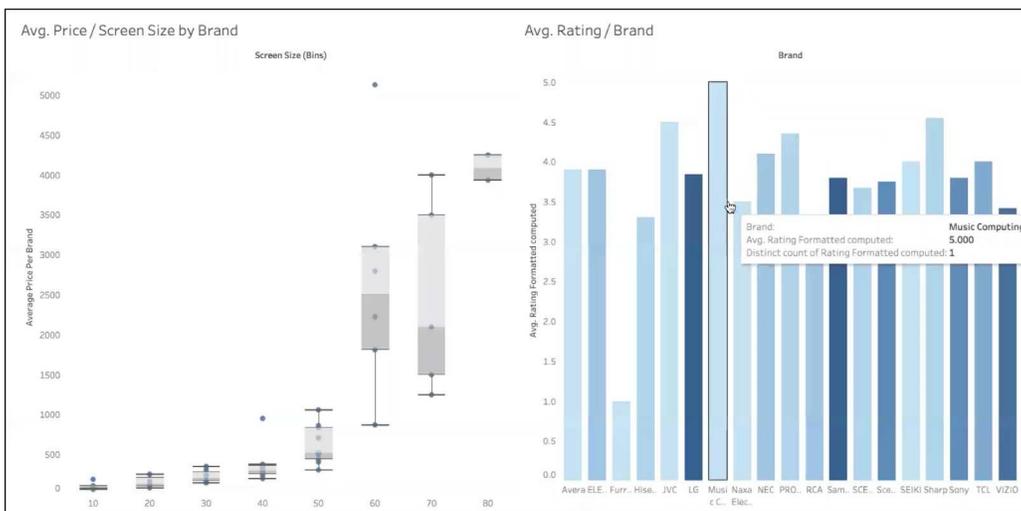
All of the data, both the original data and transformed data, is stored in Import.io's cloud-based service, so you can access it from any device via a browser.



Example of using Import.io Transform to remove extra words from a customer ratings column. This enables you to visualize and compare consistent data.

Web Data Insights

Data visualization takes the abstract and turns it into something more tangible. Import.io Insights helps users gain valuable intelligence about their market, products, competitors, consumer sentiment, etc. by visualizing data in charts and graphs. With Insights, you can better detect outliers, possible errors, patterns, and may even spot future trends. For instance, correlate the customer sentiment ratings above with brands and prices by visualizing this data in a chart, such as the one below.



Maximizing the Value of Web Data

Or, create complex dashboards comparing pricing and other attributes like the one below.



Import.io Insights are stored in the Import.io cloud and available to access from any device and share with others. Web data visualization provides clarity when simply looking at numbers or rows of data is not enough.

SOLUTION ADVANTAGES

By employing the Import.io solution, organizations realize several advantages:

- **Ease of use** - With Import.io, you can easily extract almost any web data. You can schedule automated extractions that update the transformed data and ensure disparate data elements conform to a common format. Ultimately, this means you can move past the mechanics of extraction and transformation and focus on what matters: ensuring maximum insights are derived.
- **Up-to-date data** - The solution's automation means that your organization can do ongoing extraction, with minimal effort. As a result, organizations can ensure they're always using the most up-to-date data.
- **Accurate data** - Import.io enables teams to eliminate not only the effort associated with manual extraction and transformation, but also the potential mistakes associated with manual efforts as well.
- **Scalability** - The Import.io platform eliminates the bottlenecks associated with manual, labor-intensive extractions, and it offers the scalability to support extractions of millions of data points and web queries.

Maximizing the Value of Web Data

CONCLUSION

The strategic benefits of web data are already being realized by virtually all Fortune 5,000 companies in amazing ways. With Import.io, you can get past the challenges of extracting the data, while focusing on making the data useful with Transform and actionable with Insights.

“Import.io offers sophisticated Web extraction capabilities that are very easy to use. Combined with Import.io, we are able to leverage powerful machine learning and data science algorithms to automatically extract and analyze data at scale.”

FiscalNote