

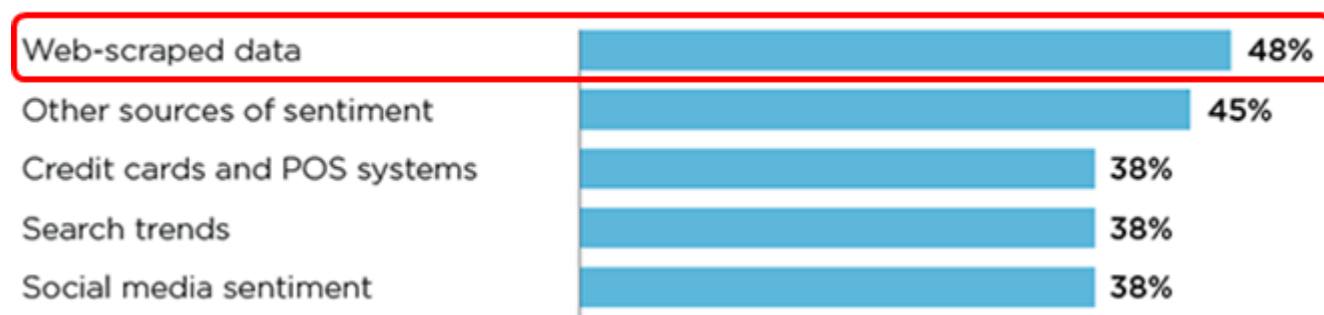
## Alternative Data in Action: Web-Scraping

January 14, 2019

In December, Nasdaq announced they intended to acquire Quandl, an alternative data company. As we wrote at the time, this represents an inflection point for the industry as [alternative data goes mainstream](#).

The early adopters of alternative data were the most sophisticated quantitative hedge funds who had the expertise and resources to take in the often unstructured data and incorporate into their investment models. Now however, usage of alternative data is expanding to more traditional asset managers. [Recent Greenwich Associates research](#) shows that 50% of institutional investors plan to increase their usage of alternative data in the coming year. Among the various types of alternative data available, web-scraped data is the most popular.

### USAGE OF ALTERNATIVE DATA SETS



Note: Based on 40 respondents. Source: [Greenwich Associates 2018 Alternative Data Customer Journey Study](#).

## Internet Insights

Web scraped data, as the name crudely implies, refers to data that has been harvested from public websites. The companies that specialize in this type of data collection write programs that access targeted websites and collect and store the scraped information on a periodic basis. In some cases vendors will use public APIs as a way to access the data within those pages directly without visiting the actual website. Vendors in this space include Quandl, Savvr, Thinknum, and Yipit.

With 4 billion webpages and 1.2 million terabytes of data on the internet there is a mountain of information that can be valuable to investors. Types of web scraped data include:

- **Job Listings:** A company that is increasing hiring and headcount is likely experiencing growth.
- **Company Ratings:** Sites like Glassdoor allows employees to rate their company; increasing ratings, especially (in conjunction with increasing job listings) can be another growth indicator.
- **Online Retail Data:** High product rankings on online retailers suggest strong sales for those product

manufacturers. On the flip side, heavy discounting of products suggest weak sales.

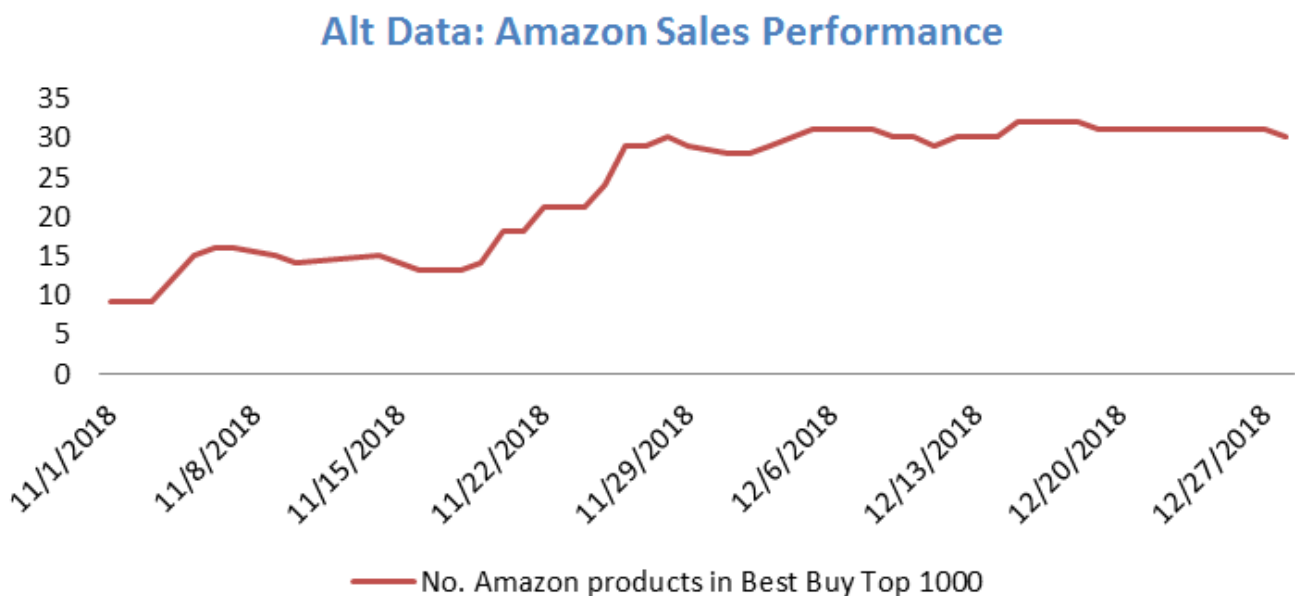
## Example: The Information Advantage

On December 26th Amazon (\$AMZN) [announced that they had had a record holiday season](#), including:

- **"Echo Dot was the #1 top-selling product across all categories on Amazon"**
- **"Echo Dot and Fire TV Stick...were not only the top-selling Amazon devices this holiday season, but they were also the best-selling products...across all of Amazon"**
- **"Millions of Prime members voice shopped with Alexa for gifts, Amazon devices and everyday household essentials"**

Amazon's success over the holiday season clearly wasn't yet factored into the stock price which has soared by 23% since the announcement while the S&P 500 remained roughly flat.

For users of alternative data though, the strong performance of Amazon products wouldn't have necessarily been a surprise. Data below provided by Thinknum sourced through Best Buy's website shows the increasing sales strength of Amazon products through the holiday period starting from Black Friday.



Source: Thinknum

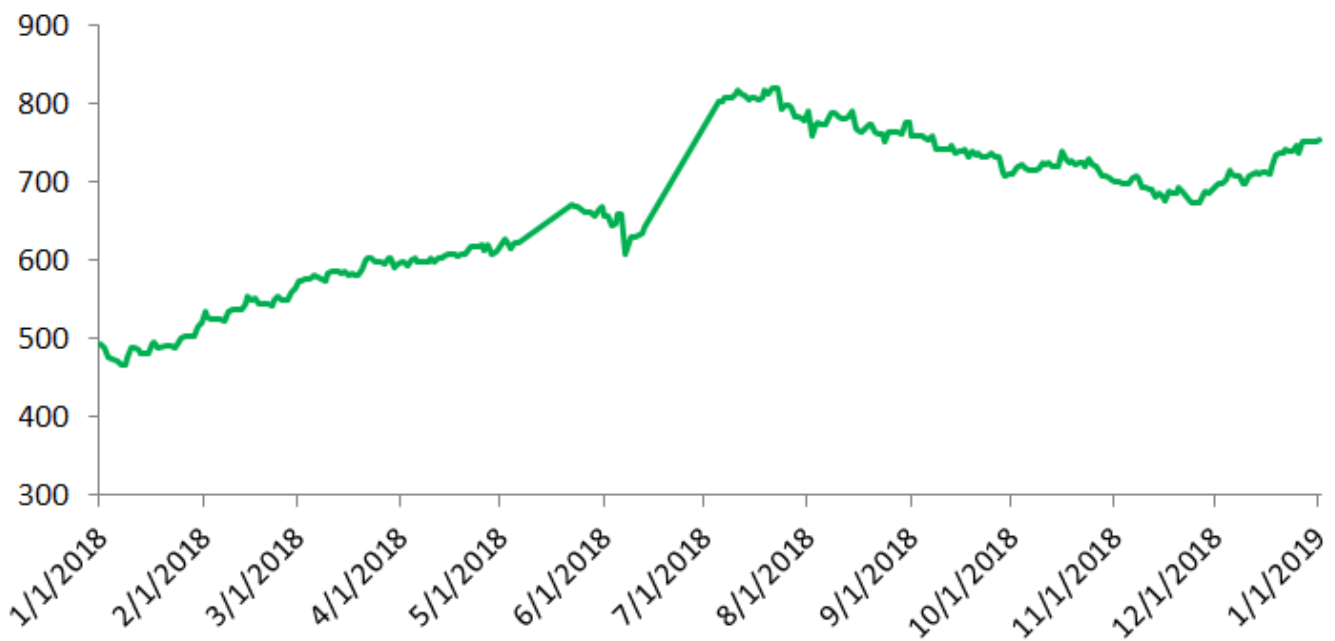
In addition, also from Best Buy's website, we can specifically see that the Echo Dot and Fire TV Stick were top sellers in their category, as noted in the official Amazon press release.

Best Buy Rank - Hardware / Devices	Product
1	Amazon - Fire TV Stick with Alexa Voice Remote - Black
2	Amazon - Echo Dot (3rd Gen) - Charcoal
3	Google - Chromecast (Latest Model) Streaming Media Player - Charcoal
4	Apple - AirPods - White
5	Nintendo - Switch 32GB Console - Neon Red/Neon Blue Joy-Con
6	Toshiba - 32" Class - LED - 720p - Smart - HDTV - Fire TV Edition
7	Beats by Dr. Dre - Powerbeats <sup>3</sup> Wireless - Black
8	Amazon - Fire TV Stick 4K with All-New Alexa Voice Remote
9	Apple - Lightning-to-3.5mm Headphone Adapter - White
10	Lenovo - 130-15AST 15.6" Laptop - AMD A6-Series
11	Amazon - Echo Dot (2nd generation) - Smart Speaker with Alexa - Black
12	Rocketfish™ - 4' 4K UltraHD/HDR In-Wall Rated HDMI Cable - Black
13	Sony - PlayStation 4 1TB Call of Duty: Black Ops 4 Console Bundle - Jet Black
14	Google - Home Mini - Smart Speaker with Google Assistant - Chalk
15	Samsung - 43" Class - LED - NU6900 Series - 2160p - Smart - 4K UHD TV with HDR

Source: Greenwich Associates and Thinknum, Dates 12/20/2018 - 12/25/2018.

And the importance of Alexa, Amazon’s AI powered, voice-controlled digital assistant, could have been seen months earlier by looking at data for ‘Amazon Alexa’ job postings on Amazon’s corporate website. Over the course of 2018, the number of open positions related to Alexa increased by 53% from around 500 to over 750.

### Alt Data: "Amazon Alexa" Job Listings



Source: Thinknum

Of course, alternative web-scraped data doesn’t have all the answers. While it may indicate which products

are selling well, it doesn't quantify the impact that may have on a large diversified company like Amazon. And although an increase in job listings is an indicator of growth, it could also mean an increase in cost which could impact profitability.

However, it is clear that alternative data such as web scraped data can provide important new information about a company's business and outlook. Sometimes this data has value on its own, but more often true value is derived when combined with other data sources, both traditional and alternative, and qualitative analysis.

---

[www.greenwich.com](http://www.greenwich.com) | [ContactUs@greenwich.com](mailto:ContactUs@greenwich.com)

Coalition Greenwich, a division of CRISIL, an S&P Global Company, is a leading global provider of strategic benchmarking, analytics and insights to the financial services industry.

We specialize in providing unique, high-value and actionable information to help our clients improve their business performance.

Our suite of analytics and insights encompass all key performance metrics and drivers: market share, revenue performance, client relationship share and quality, operational excellence, return on equity, behavioral drivers, and industry evolution.

## About CRISIL

CRISIL is a leading, agile and innovative global analytics company driven by its mission of making markets function better. It is majority owned by S&P Global Inc., a leading provider of transparent and independent ratings, benchmarks, analytics, and data to the capital and commodity markets worldwide.

CRISIL is India's foremost provider of ratings, data, research, analytics, and solutions with a strong record of growth, culture of innovation, and global footprint.

It has delivered independent opinions, actionable insights and efficient solutions to over 100,000 customers through businesses that operate from India, the U.S., the U.K., Argentina, Poland, China, Hong Kong, and Singapore.

For more information, visit [www.crisil.com](http://www.crisil.com)

## Disclaimer and Copyright

This Document is prepared by Coalition Greenwich, which is a part of CRISIL Ltd, an S&P Global company. All rights reserved. This Document may contain analysis of commercial data relating to revenues, productivity and headcount of financial services organisations (together with any other commercial information set out in the Document). The Document may also include statements, estimates and projections with respect to the anticipated future performance of certain companies and as to the market for those companies' products and services.

The Document does not constitute (or purport to constitute) an accurate or complete representation of past or

future activities of the businesses or companies considered in it but rather is designed to only highlight the trends. This Document is not (and does not purport to be) a comprehensive Document on the financial state of any business or company. The Document represents the views of Coalition Greenwich as on the date of the Document and Coalition Greenwich has no obligation to update or change it in the light of new or additional information or changed circumstances after submission of the Document.

This Document is not (and does not purport to be) a credit assessment or investment advice and should not form basis of any lending, investment or credit decision. This Document does not constitute nor form part of an offer or invitation to subscribe for, underwrite or purchase securities in any company. Nor should this Document, or any part of it, form the basis to be relied upon in any way in connection with any contract relating to any securities. The Document is not an investment analysis or research and is not subject to regulatory or legal obligations on the production of, or content of, investment analysis or research.

The data in this Document may reflect the views reported to Coalition Greenwich by the research participants. Interviewees may be asked about their use of and demand for financial products and services and about investment practices in relevant financial markets. Coalition Greenwich compiles the data received, conducts statistical analysis and reviews for presentation purposes to produce the final results.

THE DOCUMENT IS COMPILED FROM SOURCES COALITION GREENWICH BELIEVES TO BE RELIABLE. COALITION GREENWICH DISCLAIMS ALL REPRESENTATIONS OR WARRANTIES, EXPRESSED OR IMPLIED, WITH RESPECT TO THIS DOCUMENT, INCLUDING AS TO THE VALIDITY, ACCURACY, REASONABLENESS OR COMPLETENESS OF THE INFORMATION, STATEMENTS, ASSESSMENTS, ESTIMATES AND PROJECTIONS, ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE ARISING OUT OF THE USE OF ALL OR ANY OF THIS DOCUMENT. COALITION GREENWICH ACCEPTS NO LIABILITY WHATSOEVER FOR ANY DIRECT, INDIRECT OR CONSEQUENTIAL LOSS OR DAMAGE OF ANY KIND ARISING OUT OF THE USE OF ALL OR ANY OF THIS DOCUMENT.

Coalition Greenwich is a part of CRISIL Ltd, an S&P Global company. ©2024 CRISIL Ltd. All rights reserved.