Intelligence Node

Product Matching & Data Scraping



High-Velocity Scraping

Share of Search

Shipping Cost

Attribute Standardization

Price Monitoring

Proprietary Product Matching with Unbeatable Accuracy (SLA Guarantee)

As a brand or a retailer, you want to ensure that any product matching you are doing is providing precise visibility into the comprehensive set of substitutes that influence your shoppers' purchase choice. Today's shoppers are making decisions based on a multitude of dynamic factors that can change in real-time. Leading brands and retailers digest this context and inform their go-to-market strategies accordingly. Accurate, reliable, and timely data is a foundational infrastructure for any AI-powered or agentic system in consumer retail today.

Intelligence Node's product matching solution is powered by 'Sherlock AI' (proprietary AI technology) and our patented 'similarity engine' which makes it highly intuitive, **99%** accurate, fast, and scalable. It uses a **three-pronged approach** to product matching which enables it to match products despite missing UPC codes, inaccurate descriptions, varied nomenclature, and insufficient product descriptions:

- **Key Product Attribute Comparison:** We use product specific attributes for comparing products to ensure the match is valid.

- **Image Comparison:** We use image similarity (using CNN vector embeddings) to judge if a product-candidate pair looks visually similar.

- **Text Comparison:** We use two methods to achieve this - the first method is comparing word for word to identify an exact comparison, and the second method is using a custom-tuned BERT for capturing semantic similarity between the product texts.

Copyright © Intelligence Node. All rights reserved.

Using the above three-pronged approach ensures that all matches are vetted multiple times by looking at every aspect of the match. And in the few cases where the results derived by algorithms have a low probability, the matches are evaluated by data analysts to ensure a **99%** accuracy level every time (it is written in our contractual SLAs). We further standardize the attribute nomenclature, identify miscategorized products, and ensure our matches are accurate from both 'recall' and 'precision' perspectives.

With Intelligence Node, retailers and brands can balance the competitiveness of prices for shoppers with margin guardrails that drive business sustainability and performance. Our methods focus on one-to-one identical matching for branded products and the discovery of similar products for private labels with one-to-one similar matching. Moreover, Intelligence Node can tailor its matching algorithm to meet your internal criteria for tracking similar products in your category.

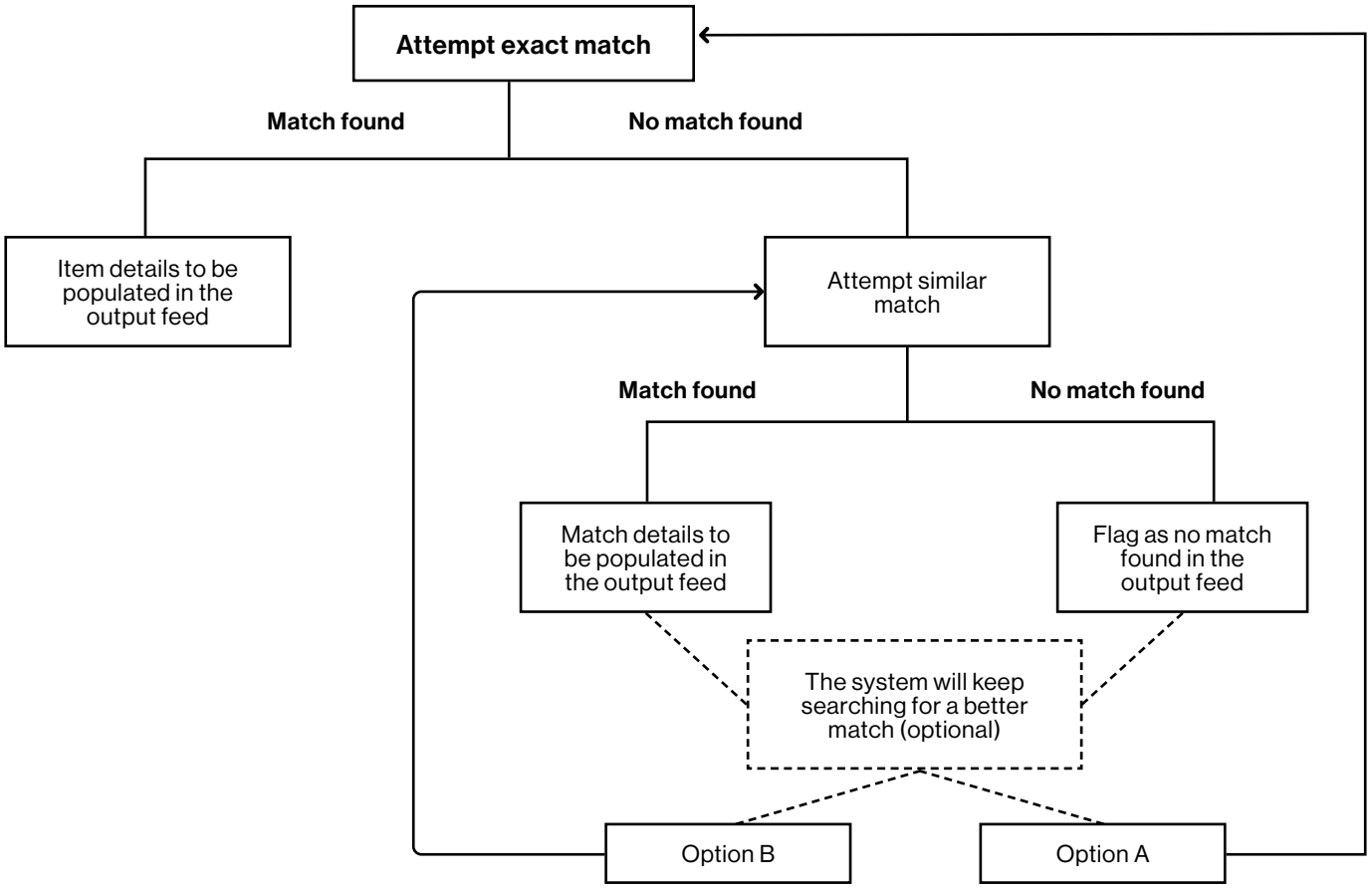
Exact match	Variant match	Similar match
Like to like products from the same brand	Same product from the same brand with only variation/difference in pack size (pack of 1 vs pack of 3), quantity (30 ml vs 150 ml) or both. Intelligence Node calculates unit-based pricing to ensure true benchmarking	There are 2 approaches that are followed for flagging a similar match - • Attribute driven matches • Shopper driven matches Both the approaches have been explained in detail in the following slides

Copyright © Intelligence Node. All rights reserved.

A Dynamic Approach to Product Matching

At Intelligence Node, we are always striving to enhance our product offerings by leveraging Artificial Intelligence and advanced, predictive analytics. Instead of finding one time matches for our customers, we are now applying a dynamic approach to product matching where we continuously crawl competitor websites to find better, closer, and the most relevant matches for your products even in cases where there is already a similar or variant match identified in the first instance.

The below workflow breaks down the working of the dynamic pricing approach to Product Matching:



Copyright © Intelligence Node. All rights reserved.

Real-time High Velocity eCommerce Data Scraping

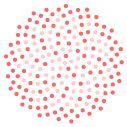
Real-time data 'listening' or data scraping refers to our process to capture price and other product details from competitive domains in real-time by hitting a product URL. These details are not previously stored in Intelligence Node's database and represent what happens on the website at that specific moment.

Many data providers use a process of 1-by-1 requests to collect data, which makes it difficult to scale and maintain high data accuracy levels. On the other hand, our system makes multiple data collection calls in parallel. The result is high velocity, highly accurate information that requires minimal human oversight. We make it easy for our clients to access their competitive domains in real-time via a standard API (Application programming interface), where the client just needs to input the product URL. Machine learning-enabled, high velocity data listening helps you spot opportunities and market trends by reviewing competitors' catalogs and price movements in real-time.

Below we discuss seven different methods of data mining that deliver high velocity, volume, and variety:

1. Full domain data listening
2. Flexible coverage of specific categories
3. Identification of key product details from the product page
4. Tracking the final price of the product
5. Capture stock level from multiple sources
6. ZIP code level data scraping
7. 1P vs. 3P seller parsing

Copyright © Intelligence Node. All rights reserved.



1. Full domain data listening

Full domain data listening involves collecting all intelligence available on an eCommerce site down to the product description level. The ability to accurately and comprehensively collect information from an eCommerce site is challenging as websites layouts are dynamic, and there are multiple layout/responses across products to capture. Moreover, no two websites use the same structure and nomenclature, making it an ongoing and continuously evolving process.

To capture full domain data, we first scan an entire website to capture all URLs, followed by the use of an NLP (Natural Language Processing) based AI algorithm. - The NLP delivers classification using **LSTM** (Long Short Term Memory) on **BERT** (Bidirectional Encoder Representations from Transformers embeddings) to classify the URLs into the product, listing page URLs, and separate out invalid URLs, which helps us narrow down the scope and save on cost and time.

Intelligence Node's proprietary algorithms are designed and triggered using the same techniques as used by Google, thereby enabling our technology to mimic Google search behavior. As a result, we are able to collect the same accurate and precise information that Google retrieves. However, our machine learning capabilities allow us to get to a deeper level of granularity for consumer retail specific use cases.

Intelligence Node can scrape the full domain and provide information in a predefined format, e.g. Text/CSV, JSON, XML, etc.

2. Flexible coverage of specific categories

If you do not want to opt for a full domain data capture, our high velocity data scraping approach provides the **flexibility to choose specific categories based on your business needs, thereby reducing cost**.

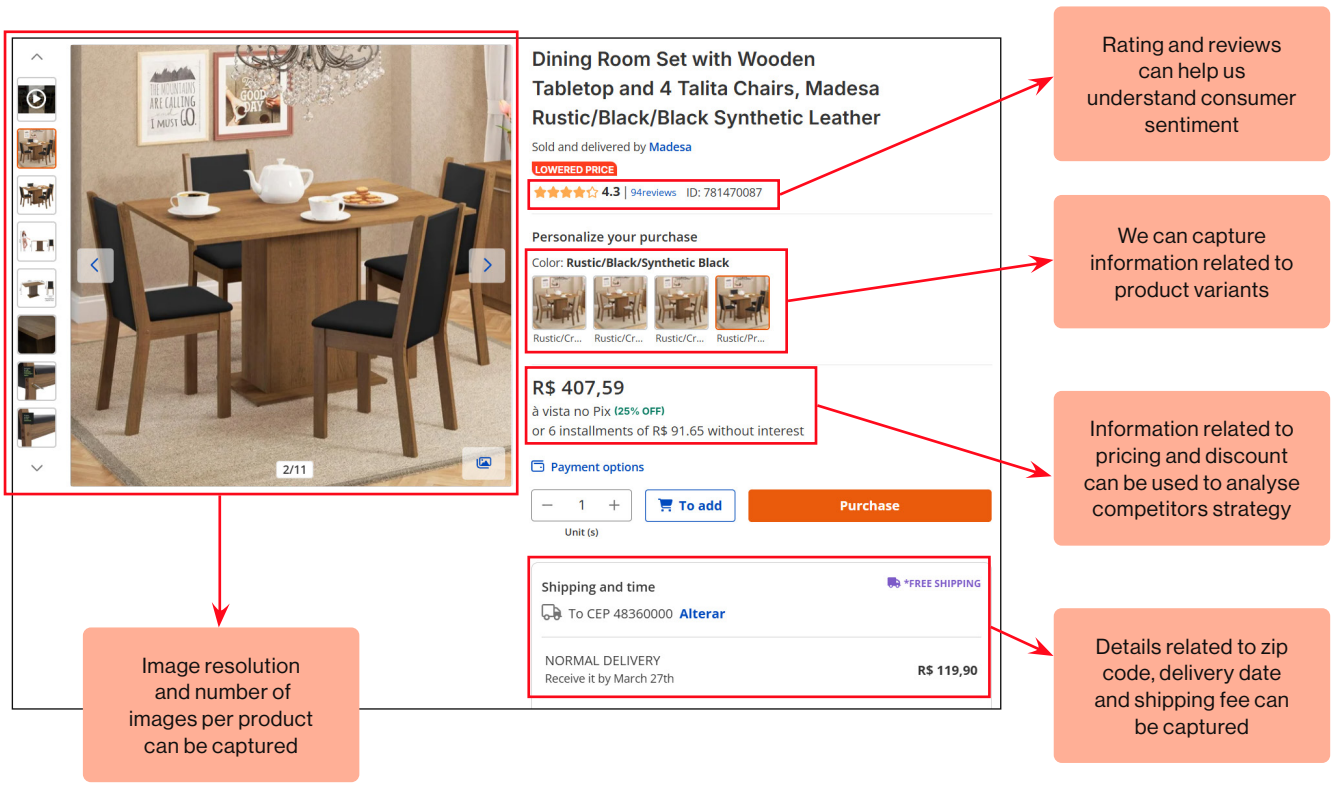
Copyright © Intelligence Node. All rights reserved.

3. Identification of key product details from the product page

High velocity scraping captures key product details, such as

- product variants,
- ratings and reviews,
- shipping costs and time,
- pricing information, etc.

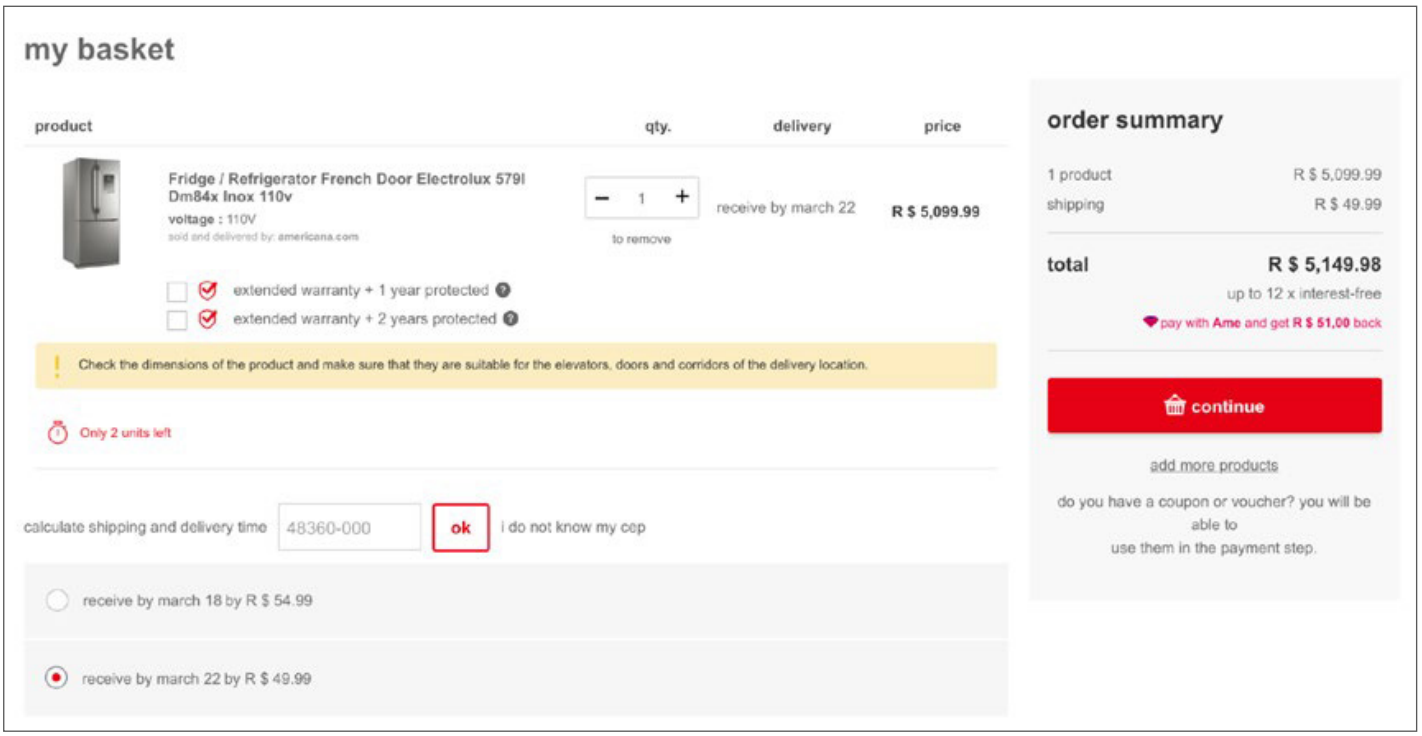
from the product description page. This information helps businesses decode competitor strategy and make product related decisions accordingly. We render JavaScript to capture items that are visible only on runtime. Additionally, we use headless browsing to load web pages and perform automated actions (such as selecting sub menus) in a similar way that a human would. In the case of the CAPTCHA test, we resolve this with human-assisted analysis and automated CAPTCHA breaking techniques.



Copyright © Intelligence Node. All rights reserved.

4. Tracking the final price of the product

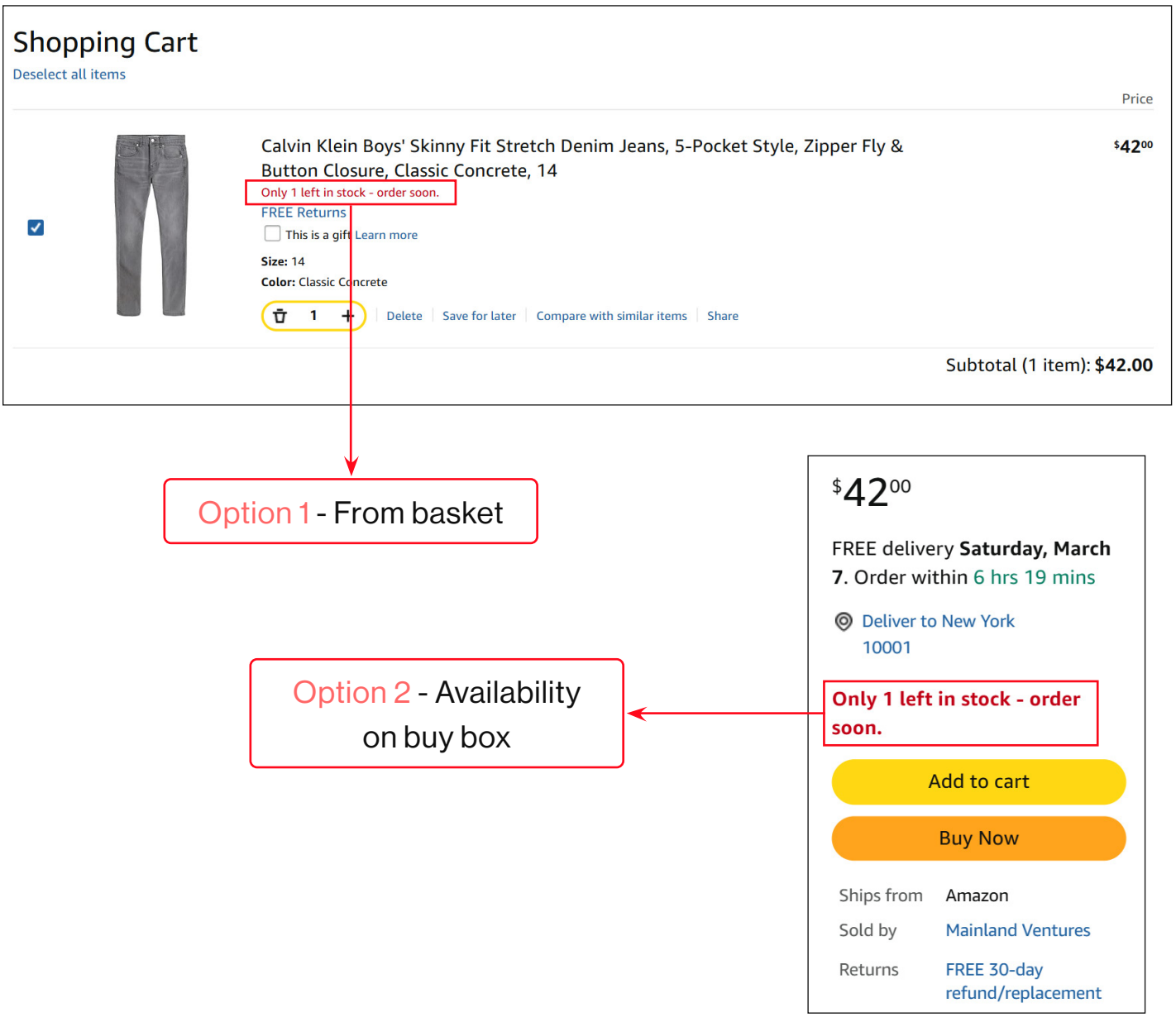
Intelligence Node's sophisticated eCommerce data scraping capabilities help our customers obtain the product's final price which is only visible in the basket and otherwise not easily accessible. The final price includes shipping fees, taxes, duties, and any other charges that may not be visible on the product description page.



Copyright © Intelligence Node. All rights reserved.

5. Capture stock level from multiple sources

Our “smart recipe scraping” algorithms can **track sales velocity by scraping inventory count from multiple sources** including product pages, shopping carts, the Amazon buy box, etc., We analyze the frequency at which the inventory levels are depleting over a certain period.

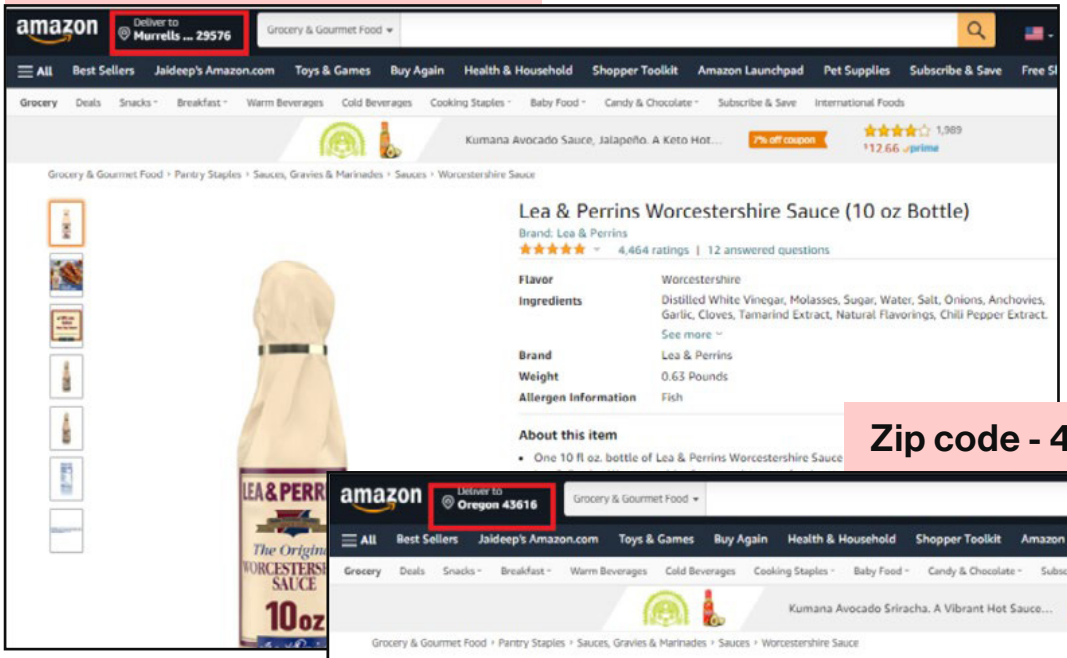


Copyright © Intelligence Node. All rights reserved.

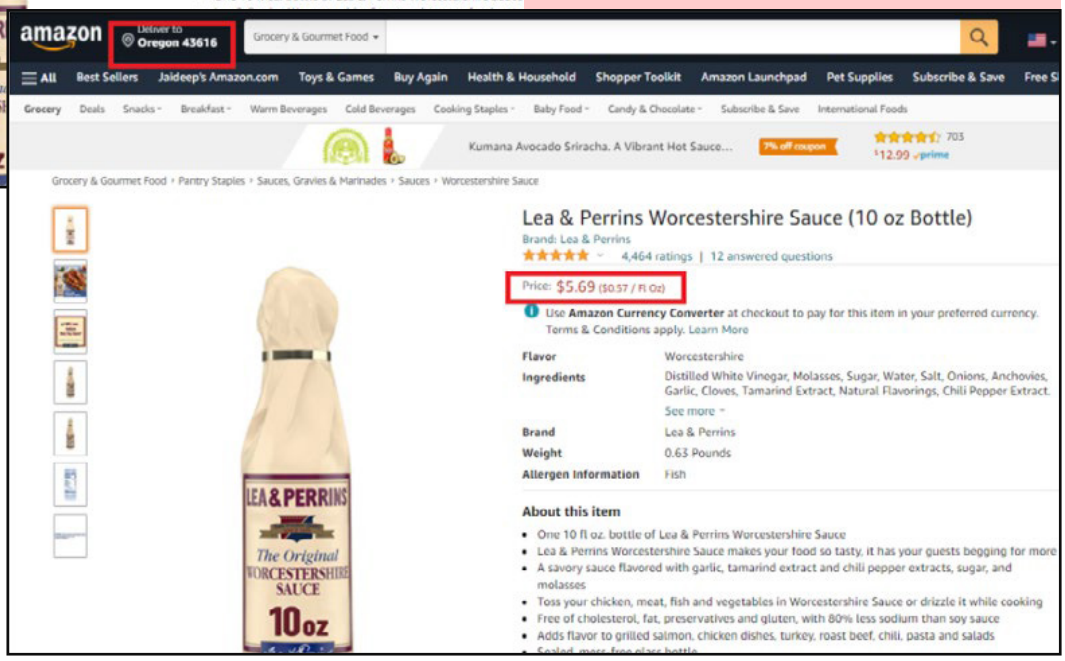
6. ZIP code level data scraping

Prices can differ largely from one ZIP code to another depending on demand, shopper demographics, supply, and geographic placement. Hence, finding a data provider that can parse data at the ZIP code level can help identify trends and benchmark against competition at a more granular level. Intelligence Node's patented machine learning algorithms can compare identical, private labels, or similar products down to the ZIP code level. These algorithms can help brands and retailers identify and compare differences in prices and store descriptions at different locations.

Zip code - 29576 (11.89 USD)



Zip code - 43616 (5.69 USD)



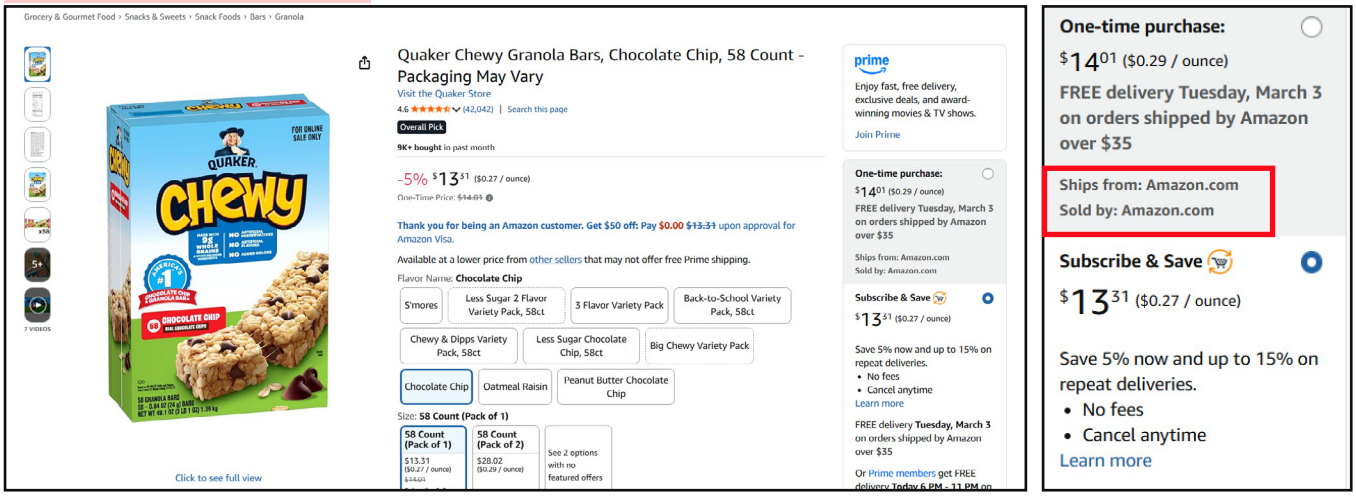
Copyright © Intelligence Node. All rights reserved.

7. 1P vs. 3P seller parsing

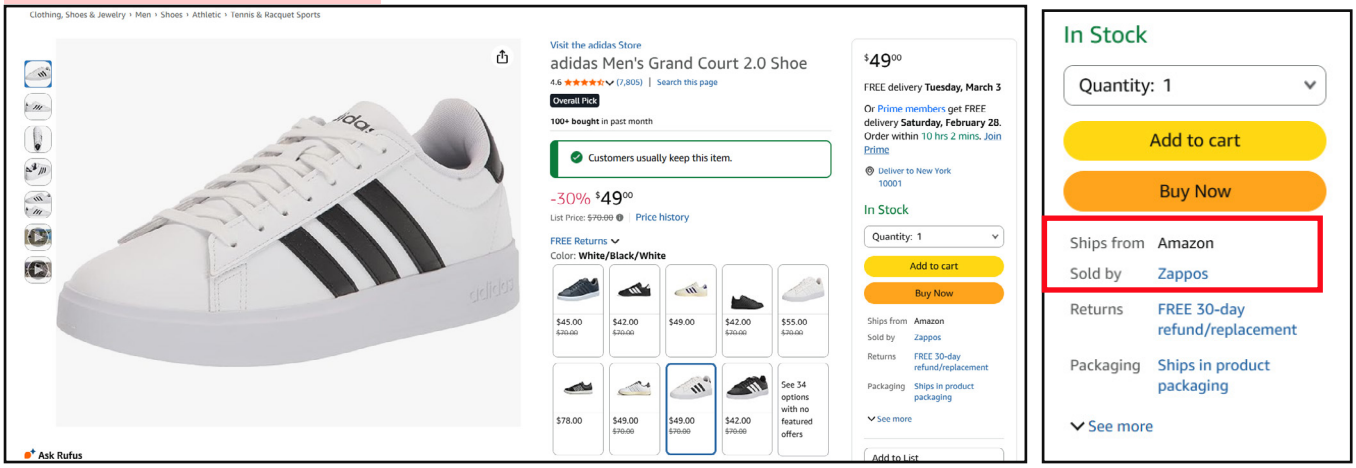
Intelligence Node parses 1P (1st party) and 3P (3rd party) seller intelligence to identify sellers which are relevant for marketplaces such as **Amazon**, **Walmart**, and **Kroger** in the near future.

Knowing whether the seller is 1P or 3P on a marketplace platform is critical for a retailer to appropriately monitor the competition and take appropriate action in terms of price as well as logistics related to order fulfillment. It also enables a retailer to accurately benchmark its assortment against various sellers on marketplace platforms.

Example: 1P seller



Example: 3P seller



Copyright © Intelligence Node. All rights reserved.

About Intelligence Node

Intelligence Node is a real-time product intelligence data powerhouse powering modern retail. We sit at the center of the digital shelf, transforming pricing, assortment, promotions, search, content, and media signals into executable intelligence that drives growth, precision, and accountability.

Our proprietary product graph and industry-leading product matching methodology, delivering **99%** accuracy, tracks over **1.2 billion** products, **190,000+** brands, and **1,400+** categories. This intelligence fuels more than $600 billion in global retail revenue, giving retailers unmatched visibility into how products perform, compete, and convert across every channel.

Acquired by Interpublic (NYSE: IPG) in 2024, Intelligence Node is now integrating with Acxiom and Omnicom Media (Omnicom, NYSE: OMC), serving as a foundational technology at the heart of the world's largest advertising, sales, and marketing ecosystem. Together, we are changing how retail performance, media strategy, and creativity converge in the age of intelligent commerce.

Drive Sales Growth With Real-time Data Scraping & Matching

Explore how Intelligence Node's proprietary data scraping and product matching technology can help you drive conversions, profitability, and customer satisfaction.

BOOK A CUSTOM DEMO