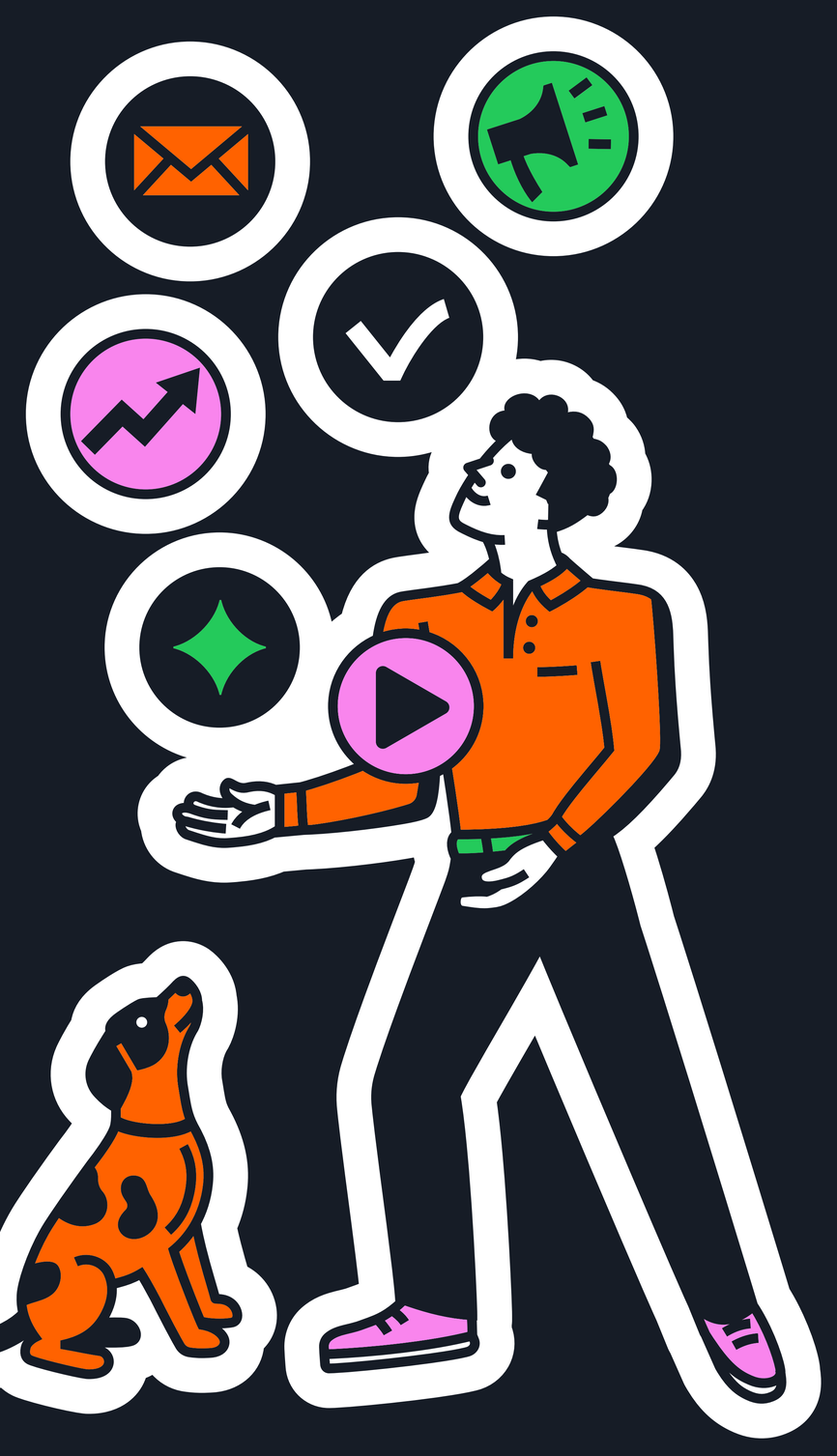
amazon ads

The New Standard: How Benchmarks Are Transforming Performance Evaluation



1 Introduction

Page 03

2 Benchmark Intelligence in Action: From Data to Insights

Page 04

3 Technical Architecture

Page 05

4 The Customer Impact

Page 10

5 Looking Ahead

Page 11

Introduction

The Benchmarking Challenge

Performance uncertainty erodes advertiser confidence and ROI. Without reliable benchmarks, advertisers struggle to answer fundamental questions: How should I allocate my advertising budget? What performance should I expect in my category? Are my campaigns performing well? How do I compare to similar brands? Where should I optimize my advertising strategy?

Many advertisers rely on ad hoc analysis rather than accessing ongoing, trended benchmark data over time, creating gaps in their ability to track performance evolution and identify emerging patterns. As Forrester notes, **“The real value of benchmarks lies in understanding them in the context of industry dynamics and using them to diagnose underlying issues in the organization.”**

Marketers need insights on what 'good' looks like to guide investment decisions and measure performance. Amazon Ads benchmarks deliver this through a data-driven approach built on peer matching, historical analysis, and performance normalization. By contextualizing campaign performance against peers and competitive dynamics, benchmarks transform backward-looking insights into a forward-looking compass for confident decision-making.

Amazon Ads benchmarks provides performance context across the full funnel, enabling advertisers to:



Make confident investment decisions:
Evaluate new opportunities with category-specific context, reducing risk in unfamiliar ad products and identifying optimization priorities—whether amplifying outperforming areas or addressing performance gaps relative to peers.

Accelerate business growth through data-driven planning:
Inform annual budget allocation, evaluate media mix effectiveness, and identify growth opportunities with clear performance standards across ad formats and campaign types.

Enhance workflow efficiency through streamlined analysis:
Access benchmarks via Ad Console reporting or integrate via APIs and tools that advertisers already use for standardized reporting, and scaled analysis.

This white paper explains the data-driven methodology behind reliable benchmarks, which evaluates brand performance against peers across ad formats, campaign goals, time intervals, and categories to ensure relevance.

Benchmark Intelligence in Action: From Data to Insights

Foundations of Benchmarking

Effective benchmarking requires balancing competing priorities: statistical rigor, competitive relevance, and data protection. The best methodologies address this through three core principles.

Data Aggregation and Privacy
Advertiser data protection is central to our benchmarking approach. Reliable benchmarks demand sufficient sample sizes to achieve statistical rigor, yet individual advertiser data must remain confidential. Quality benchmarks aggregate performance data across peer cohorts large enough to ensure robust calculations while maintaining strict confidentiality protections. No single advertiser's performance is identifiable.

Peer Construction
Meaningful benchmarks require accurate competitive grouping that reflects current dynamics. Rather than relying on categorical classifications alone, a reliable benchmark should identify empirically comparable competitors by analyzing brand characteristics, product portfolios, and positioning. For example, at the core of Amazon Ads benchmarks is PeerSets, a system that identifies direct competitor brands for any given brand within a category context. For each focus brand—defined as the advertiser brand for which benchmark metrics are computed—PeerSets surfaces a set of peer brands that genuinely represent the advertiser's competitive landscape. This rigorous matching approach ensures peers reflect genuine competitive dynamics and provide actionable insights.

Contextual Relevance
Fair performance comparisons must account for current dynamics and campaign characteristics that influence outcomes. Normalizing benchmarks across multiple dimensions enables valid apples-to-apples analysis. This contextual approach recognizes that performance expectations vary by market segments, product category, and campaign type.

Multi-Dimensional Performance Analysis

Benchmarks provides performance context across multiple critical dimensions:



Category Specificity: Brand category-level analysis ensures comparison against genuinely competitive peer sets



Ad Product Type: Coverage across ad products. For example: Sponsored Products, Sponsored Brands, Sponsored Display, Amazon DSP, and Sponsored TV



Ad Format: Streaming TV, online-video, display, and audio



Campaign Objectives: Segmentation by awareness, consideration, and conversion goals



Time Intervals: Daily, weekly, and monthly granularity to support varied planning cycles



Geographic and Marketplace Context: Country, marketplace, and regional dynamics



Product Pricing: Average sales price to account for category-specific conditions

The challenges of Benchmark calculation

Traditional benchmarking approaches face several key challenges:

1 **Identifying Truly Comparable Peers:**
Simple category classifications fail to capture the nuanced competitive landscape. Brands vary significantly in product portfolios, positioning, and price ranges—even within the same category.

2 **Ensuring Statistical Reliability While Protecting Privacy:**
Benchmarks require sufficient data to be actionable, yet revealing individual competitor information violates confidentiality. Additionally, advertisers need more than a single median comparison—they want to understand where their performance falls across the full distribution (below, at, or above average, and across percentiles).

3 **Preventing Distortion from Outliers:**
A single dominant advertiser can skew averages, minimums, or maximum performance metrics, making benchmarks unrepresentative of current dynamics.

How we compute benchmarks

Our benchmark computation methodology addresses these challenges through a **four-step process** that balances statistical rigor with category relevance while maintaining strict confidentiality protections.

Step 1: Intelligent Peer Brand Identification

PeerSets uses a hybrid approach that combines two complementary methods: a K-Nearest Neighbors (KNN) statistical model and a large language model (LLM). Each method brings distinct strengths to competitor identification, and together they produce a more complete and accurate picture of the competitive landscape than either could achieve alone.



The KNN model identifies competitor brands by calculating the mathematical distance between brands based on their brand-level revenue and average sales price within a category. Brands with similar sales volumes and price points are considered "neighbors" and therefore potential competitors. This quantitative approach is fast, scalable, and grounded in objective data—making it well-suited for identifying competitors of similar size and price positioning.

Technical Architecture



However, a purely statistical approach has limitations. The KNN model operates on numerical proximity alone and cannot capture the dynamic context needed to identify relevant smaller competitors for larger, diversified brands that span multiple product categories. This is where the large language model (LLM) complements the approach. The LLM brings semantic understanding to competitor identification by evaluating brands based on direct competition dynamics, market segment overlap, and customer base similarities. It reasons about brands the way an analyst would—recognizing that two brands can be direct competitors even if their sales volumes differ significantly, as long as they target the same customers and occupy the same segment space. This makes it particularly effective at surfacing relevant peers across varying brand sizes and identifying competitors a purely statistical model might overlook.

By combining both approaches, PeerSets achieves breadth and precision: the statistical model anchors peer identification in real market segment data, while the LLM ensures that competitive relationships are understood in the way advertisers and analysts think about them.

Step 2: Campaign Qualification and Filtering

We apply consistent eligibility criteria to ensure benchmarks reflect relevant, high-quality data. Campaigns must meet thresholds for valid operational status, minimum activity levels, and matching advertising attributes (ad format, campaign goal, product type). By segmenting benchmarks by campaign objective, we ensure comparisons reflect relevant KPIs—recognizing that success metrics differ between awareness, consideration, and conversion-focused initiatives. This approach preserves the vast majority of campaigns and ad spend while excluding data that doesn't meet quality standards, ensuring benchmarks are representative of peer performance.

Step 3: Brand-Level Aggregation

Campaign performance metrics are aggregated at the peer brand level using sum-over-sum calculations. For example, for each peer brand *i*, we calculate Click-Through Rate (CTR) across all valid campaigns. This approach reduces bias from small, fragmented campaigns and reflects true marketing effectiveness. These brand-level metrics then form the basis for spend-weighted percentile benchmarks, where each brand's influence is proportional to its ad spend and subject to confidentiality safeguards.

$$CTR_i = \frac{\sum \text{Clicks}_i}{\sum \text{Impressions}_i}$$

Step 4: Privacy-Preserving Guardrails

We employ multiple safeguards to ensure benchmark reliability while protecting advertiser confidentiality. These guardrails work together to prevent any single advertiser from being identifiable within benchmark calculations.

First, we require a minimum sample size of at least five peer brands to construct any benchmark. This threshold ensures statistical robustness while making it impossible to reverse-engineer individual advertiser performance from published benchmarks. If insufficient peers exist, no benchmark is surfaced rather than risk compromising confidentiality.

Second, we apply spend weighting with a cap to prevent dominant advertisers from distorting benchmarks. While we weight brands by their ad spend to reflect real competitive dynamics, we cap any single brand's contribution at no more than 49% of the total. This prevents a dominant advertiser's performance from overwhelming the benchmark and becoming identifiable. We then proportionally redistribute the remaining shares among peer brands to maintain statistical integrity.

Together, these guardrails preserve the representativeness of smaller brands, prevent distortion from outliers, and maintain statistical robustness—all while ensuring no individual advertiser's data can be isolated or identified within our published benchmarks.

Why Spend-Weighted Metrics Matter?

Consider why ad spend weighting matters. Suppose Brand X invests $10,000 at a 1.0% CTR, Brand Y spends $1 at a 10% CTR, and Brand Z spends $1 at a 20% CTR. An unweighted median CTR across these three brands would report 10%, implying that "typical" campaign performance is 10%. This is misleading because it treats all brands equally regardless of market segment presence. **In reality, 99.98% of the total ad spend ($10,002) is associated with the 1.0% CTR.**

A more intuitive way to think about this: normalize all campaigns to $1 each. Brand X would effectively represent 10,000 identical campaigns at 1.0% CTR, alongside two additional $1 campaigns from Brands Y and Z. Ranking all 10,002 campaigns and finding the median naturally yields 1.0%—not the outlier values of 10% or 20%. This spend-weighted approach ensures benchmarks **reflect the performance representing the majority of actual advertising investment in current dynamics,** rather than being distorted by small-budget outliers.

An Illustrative Example

To illustrate the spend-weighted percentile methodology in practice, we present a simplified example with five peer brands competing within the same category. We follow four steps to compute benchmarks:

1. Rank peer brands by their performance metric values in ascending order
2. Calculate adjusted spend shares, applying the 49% spend weight cap and redistributing excess to prevent any single advertiser from skewing results
3. Determine brand spend share blocks and their midpoints, to which each brand's metric value is assigned. A spend share block is the contiguous segment of the cumulative spend axis that a brand occupies, with its width equal to the brand's adjusted spend share. Higher-spending brands claim wider segments and carry more influence in the final percentile calculation.
4. Compute percentile benchmarks using linear interpolation between block midpoints; if a percentile lands outside the first or last block, use that block's metric value directly

Let's walk through a hypothetical example of five brands in *Table 1*, sorted by CTR in ascending order.

Table 1 - Brand-Level Performance Metrics Before Spend Share Adjustment

Brand	Total clicks	Total impression	CTR	Ad-Spend ($)	Spend Share (%)
Brand A	5,000	500,000	1%	$10,000	10%
Brand B	10,000	500,000	2%	$10,000	10%
Brand C	30,000	1,000,000	3%	$60,000	60%
Brand D	20,000	500,000	4%	$10,000	10%
Brand E	25,000	500,000	5%	$10,000	10%

With the brand-level metrics established, we check if the spend weight cap needs to be applied. Since Brand C's unadjusted spend share is 60%, exceeding the 49% cap, we reduce it to 49% and redistribute the excess 11% proportionally among the remaining four brands. *Table 2* shows the adjusted spend shares, where each brand occupies a contiguous block along the cumulative spend axis with width equal to its adjusted share. Each brand's metric value is assigned to its block's midpoint.

Table 2 - Redistributed Spend Share After Applying the Spend Weight Cap

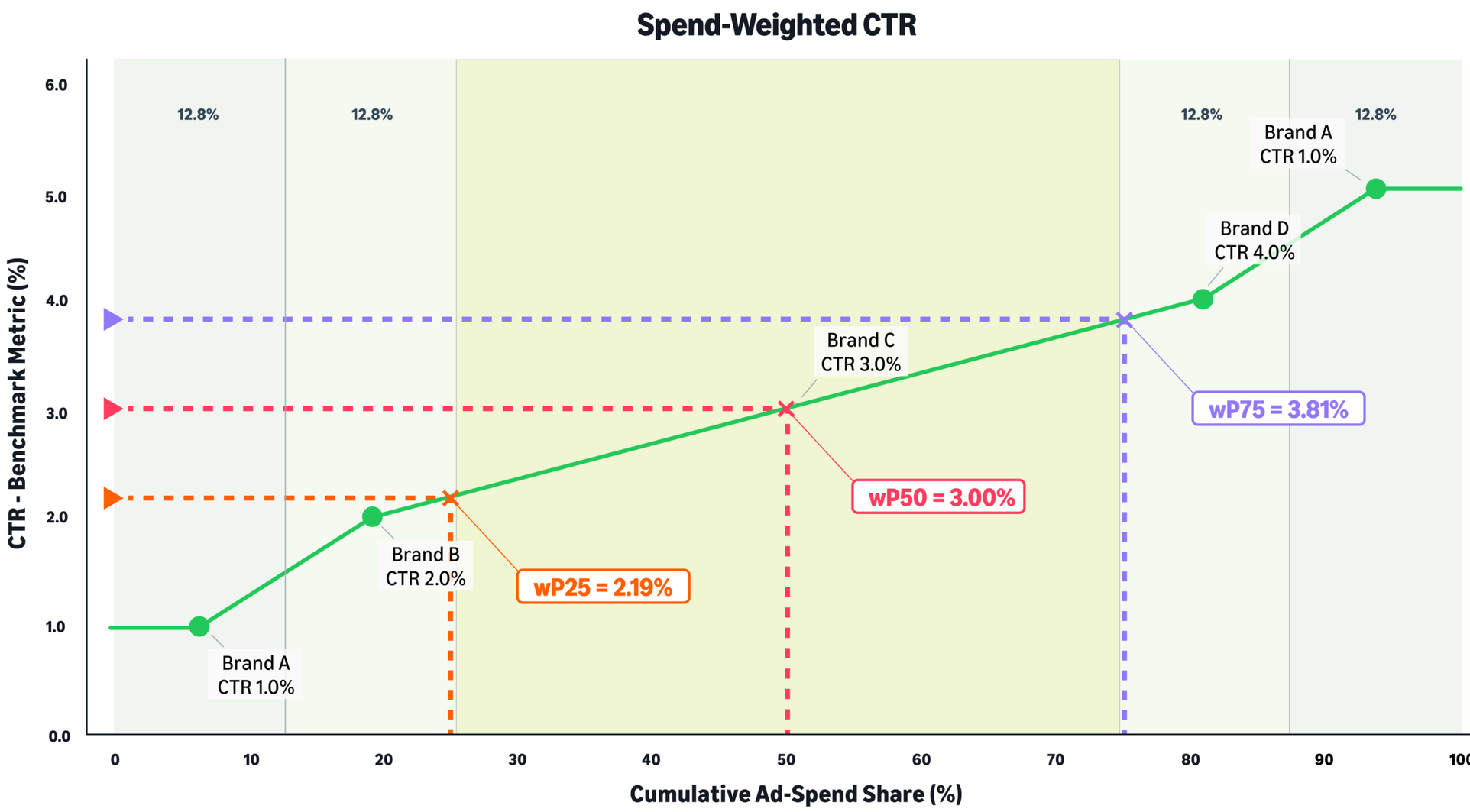
Brand	CTR %	Original Spend Share %	Adjusted Spend Share %	Cumulative Spend Share Block Start %	Cumulative Spend Share Block End %	Block Midpoint %
Brand A	1%	10%	12.75%	0.00%	12.75%	6.38%
Brand B	2%	10%	12.75%	12.75%	25.50%	19.12%
Brand C	3%	60%	49.00%	25.50%	74.50%	50.00%
Brand D	4%	10%	12.75%	74.50%	87.25%	80.88%
Brand E	5%	10%	12.75%	87.25%	100.00%	93.62%

Technical Architecture

Having determined the adjusted spend share distribution and block midpoints, we can now compute the spend-weighted percentile benchmarks through linear interpolation:

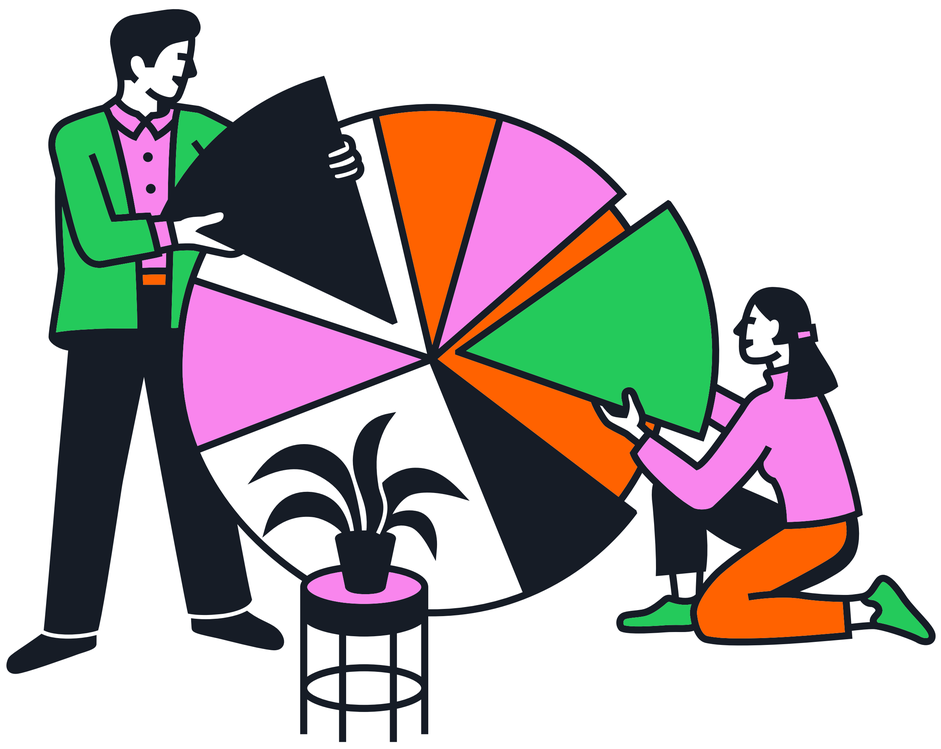
- **wP$_{50}$** (weighted median): The midpoint of Brand C's spend share block is located exactly at 50.00%, therefore wP$_{50}$ = 3.00%
- **wP$_{25}$** (weighted 25th percentile): The 25th percentile lands between the midpoints of Brand B and C's spend share blocks, so wP$_{25}$ = 2.0% + (25% − 19.12%) / (50.00% − 19.12%) × (3.0% − 2.0%) = 2.19%
- **wP$_{75}$** (weighted 75th percentile): The 75th percentile lands between the midpoints of Brand C and D's spend share blocks, so wP$_{75}$ = 3.0% + (75% − 50%) / (80.88% − 50.00%) × (4.0% − 3.0%) = 3.81%

This figure below visualizes this process. The green line represents the weighted percentile value (wP_x) at each percentile from 0 to 100, showing how wP25, wP50, and wP75 are derived.



Customer Impact

Most advertisers lack access to peer performance data and the science resources to apply the statistical rigor needed to benchmark meaningfully. Without these foundational elements, they struggle to contextualize their performance against meaningful competitive baselines, leaving them to make decisions without critical insight.



Reliable benchmarks transform this dynamic. By providing transparent, category-specific performance context, benchmarks enable advertisers to move beyond isolated metrics and make confident, informed decisions.
Advertisers use benchmarks to identify optimization opportunities, set realistic performance targets grounded in competitive dynamics, and allocate budgets more effectively across campaigns and channels. Advertisers and partners across industries have seen significant benefits from utilizing benchmarks.

“Standard media attribution shows the ‘what’ behind performance but lacks context against the market and competition. Amazon Ads’ benchmarks' closes this gap, empowering our teams to make smarter optimization and planning decisions against larger category dynamics.”

— Sandy Welsch, Executive Director, Global Commerce, WPP Media

Looking Ahead

As digital advertising evolves, advertisers increasingly demand transparent, actionable performance context. Amazon Ads benchmarks are built for this future. Our approach centers on providing comprehensive, category-specific insights through sophisticated peer matching and transparent methodology while ensuring seamless access across advertiser workflows.





By advancing innovation grounded in data quality and statistical rigor, we're helping advertisers make confident, informed decisions that drive superior business outcomes. Today we support 8 metrics: percent of purchases (new to brand), purchase rate (new to brand), cost per purchase (new to brand), click-through rate (CTR), cost per click (CPC), completion rate (video ad), cost per completed view (video ad), and cost per thousand impressions (CPM). Future enhancements will expand benchmark coverage and provide additional flexibility and levels of granularity to support advertiser success.