

High-Dimensional Analytics in E-Commerce

How to pinpoint customer cohorts that traditional averages miss

Introduction

Raw data can be like an untranslated ancient text—filled with meaning, but easily misread without the right tools. Just as scholars must understand not only the literal symbols but also the cultural context to interpret ancient writings, data analysts need to look beyond surface-level trends.

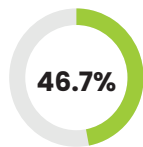
Simpson's paradox is a striking example of why context matters. Edward H. Simpson first described this phenomenon in a technical paper in 1951. It's everywhere. Aggregated data can suggest one trend, while the same data, broken into smaller, more meaningful groups, reveals the opposite. Once you are aware of the paradox, you'll never look at data the same again. The applications transcend basketball to business outcomes.

Imagine a coach analyzing shooting percentages for two fictitious players, Allen and Jordan, over a full season.

Combined Season Data:

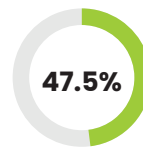
Allen:

140 shots made out of 300 attempts



Jordan:

190 shots made out of 400 attempts



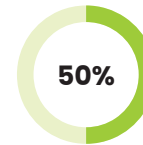
Based on this aggregate data, Jordan appears to be the better shooter overall.

However, when the coach breaks down the data by shot type (a smaller batch analysis), a different picture emerges:

Three-Point Shots:

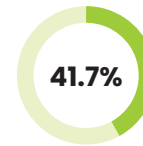
Allen:

80 made out of 160 attempts



Jordan:

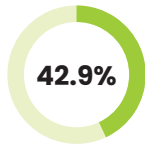
50 made out of 120 attempts



Two-Point Shots:

Allen:

60 made out of 140 attempts



Jordan:

140 made out of 280 attempts



This fine-grained analysis reveals something crucial: Allen is significantly better at three-pointers, while Jordan excels at two-pointers. The paradox occurs because Allen took a higher proportion of the more difficult three-point shots (53% of Allen's shots were three-pointers vs. only 30% of Jordan's).

The insight from examining the smaller batches leads to more strategic decisions:

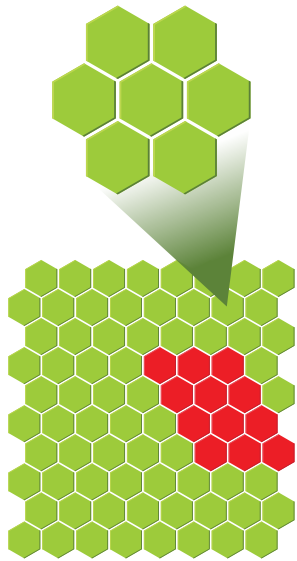
Allen should take more three-point shots.

Jordan should focus on two-point opportunities.

The team can optimize scoring by assigning shot types according to each player's strengths.

By breaking down the data into smaller batches (shot types), the coach gains actionable insights that were completely hidden in the aggregated statistics.

This is a classic example of what's known as **Simpson's paradox**—a phenomenon in which analyzing several groups of data reveals a trend that disappears or reverses when those groups are analyzed individually.



Single clusters of data reflecting few dimensions give one result. Combined clusters of data across multiple dimensions give more precise results leading to deeper insights.

Simpson's paradox can lead to serious issues with data analysis, especially when that analysis informs business decisions. In e-commerce, this paradox can distort how businesses interpret consumer behavior, campaign performance, customer satisfaction, and conversion rate optimization (CRO). But if you understand how it works and what tools can surface deep dimensionality and context-rich insights, you can avoid costly mistakes and take advantage of hidden growth opportunities that your competitors might miss.

“

“In an information-rich world, the abundance of information leads to a scarcity of attention. Information consumes the attention of its recipients.”

**—Dr. Herbert A. Simon,
1978 Nobel Prize in Economics**

To avoid Simpson's paradox and achieve high dimensionality, teams must continuously analyze behavior and experience across the many dimensions of their business—instantly and cost effectively.



Examples of Simpson's Paradox in E-Commerce

Château Digital: Revolutionizing Wine Sales Through Advanced Analytics

Château Digital, a fictional boutique online wine retailer, has transformed its customer understanding through customer analysis that goes far beyond traditional web analytics. Unlike competitors who only segment customers by a couple dimensions, like location, Château Digital analyzes purchasing behavior across hundreds of dimensions simultaneously. They can pinpoint overperforming and underperforming customer cohorts that traditional averages miss.

When analyzing browser-specific conversion patterns, they discovered customers using privacy-enabled browsers converted at nearly three times the rate of standard customers. They surfaced a previously hidden pattern of serious wine enthusiasts who valued both privacy and premium purchases. This insight allowed them to tailor marketing messages specifically to privacy-conscious consumers.

Their digital experience analytics platform revealed dramatic differences in customer journeys. Shoppers who added wines directly from homepage recommendations converted 40% higher than those browsing detailed product pages, suggesting impulse purchases often outperformed deliberate research.

Most surprisingly, customers searching for complex wine names like “Châteauneuf-du-Pape” showed different conversion patterns based on spelling accuracy, leading to improved auto-correct functionality.

By analyzing many dimensions across search terms, device types, referral sources, and sequential browsing patterns, Château Digital identified that their highest-value customers followed specific pathways through the site. This detailed behavioral insight enabled personalized experiences that increased average order values by 25% while reducing customer acquisition costs significantly.

They use high-dimensional analytics to pinpoint overperforming and underperforming cohorts—gaining insights into everything from customer churn to checkout cart abandonment.

Conversion Rate Analysis

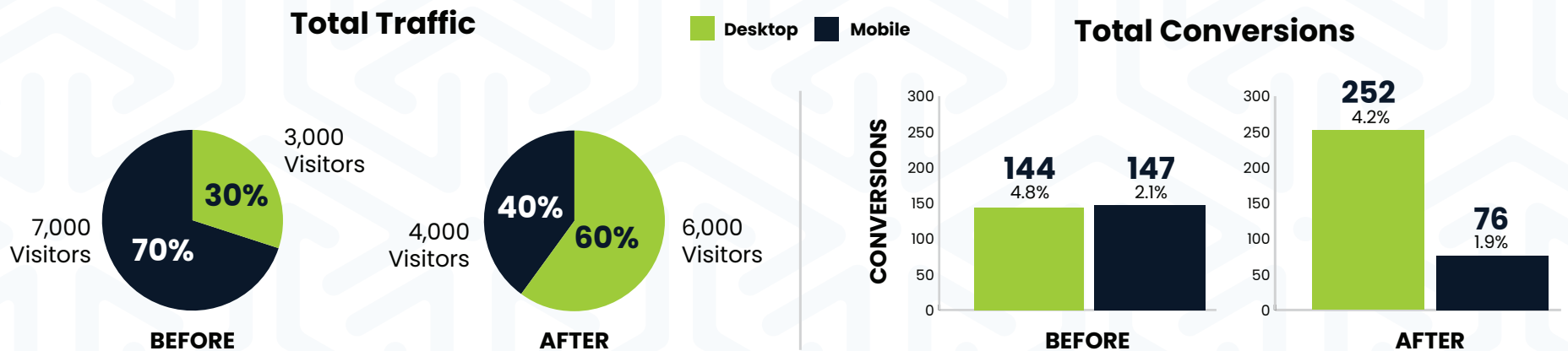
After launching a new website design, Château Digital is thrilled to see that the overall conversion rate of purchases has increased. However, when their head of e-commerce segments by traffic source, that purchase rate increase disappears.

They dig into the data and see that desktop traffic doubled after the redesign, from 30% of total traffic to 60%. Since their conversion rates are higher among desktop users than mobile users, overall conversion rates appear to improve despite both groups actually performing worse.

	Conversion before redesign	Conversion after redesign	Change
Desktop Users	4.8%	4.2%	-0.6%
Mobile Users	2.1%	1.9%	-0.2%

Imagine the website gets a total of 10,000 visitors. The charts below show how the increase in desktop traffic would affect conversions across channels and overall. The e-commerce team will want to take a look at the payment flow to understand why mobile visitors are decreasing and why their update is leading to a decrease in total mobile conversions.

The total number of conversions is higher, but not for the reasons initially suspected. Something else is driving more people to the website—and the new design does not appear to be performing well. Understanding Simpson's paradox can help you capitalize on whatever's driving traffic and make the necessary adjustments to the website to optimize purchase rates and checkout flows.



Automate Your Growth Opportunities with High-Dimensional Analytics

E-commerce leaders have likely experienced this frustrating scenario: overall purchase conversion rates look healthy, but when digging deeper, individual product categories or customer groups are actually underperforming.

Traditional product analytics tools only sample insights from two or three dimensions. Modern platforms analyze customer experience and behavior across the many dimensions, attributes, and combinations that matter to each business—for example, device type, app version, geographical region, campaign, promo code, and many more. Industry leading platforms connect behavior and experience to technical performance and business outcomes.

Customer journeys don't follow neat, linear paths. E-commerce customers rarely go straight from product page to checkout—they bounce between categories, compare options, and often abandon and return multiple times. A digital experience analytics platform must recognize these “washing machine” patterns and measure true conversion impact rather than getting fooled by a sampling of surface-level metrics.

When a tool can automatically detect that browser users are experiencing 500 errors (affecting 70 potential customers), or identify that the coupon redemption process has hidden friction points, teams are no longer flying blind. They're seeing the complete story—both the forest and the trees—which means decisions can be made based on reality, not statistical illusions.

The most dangerous checkout optimization mistakes happen when teams celebrate wins that are actually losses for critical customer groups.

Automated pattern detection reveals hidden customer journey complexities that manual analysis misses.

Real-time anomaly identification catches technical issues before they compound into significant revenue loss.

Key Takeaways for Payments Analysis

Measuring progress across flows *eventually* is not useful. When analyzing payment data, what appears as an improvement in overall payment success rates might mask declining performance across individual payment methods or customer cohorts.

For instance: Your aggregate transaction success rate might increase after implementing a new payment processor, while actually decreasing for both credit card and digital wallet payments separately. Traditional analytics tools offer limited insights into true customer experience. **Their flows and funnels show you where users dropped—not why.**

To avoid such pitfalls, always sort payment analysis by critical variables like payment method, geography, customer type, and transaction value before drawing conclusions. Watch for unexpected shifts in the distribution of payment methods, as these can create misleading aggregate trends.

Even a single milestone, like payment or checkout can have many dimensions. A payment flow will have card declines, shopping cost verification issues, synchronization issues, and app performance issues.

Fine-grained data analytics—powered by high-dimensional and high-cardinality inputs—continuously analyzes millions of attribute combinations to reveal overperforming and underperforming customer segments that traditional averages miss. With Agentic AI, these insights become autonomous actions, unlocking smarter, faster growth decisions at scale.

The Risk of Simpson's Paradox for Product Teams

What makes Simpson's paradox so troubling is how easily it can distort decision-making. Without access to all the relevant data and an understanding of what to look for, Simpson's paradox can lead to devastating outcomes for product teams:

- **Misallocating resources**

If a team draws the wrong conclusions about which user segments respond best to features, they can end up investing in the wrong areas, pursuing unprofitable customer groups, or neglecting truly high-value opportunities.

- **Abandoning successful features**

A team might shut down features that are actually beneficial for key customer groups because aggregate data shows negative overall impact. This means wasting development resources and losing potential value.

- **Doubling down on unsuccessful features**

Conversely, teams might scale up features that don't work well for most individual user groups but show positive aggregate results due to changing segment distributions, jeopardizing user experience for their core audience.

- **Conflicting metric interpretations**

Different teams analyzing the same data might reach contradictory conclusions depending on how they segment customers, leading to confusion and slower decision-making.

- **Undermining stakeholder trust**

When executives see metrics that contradict segment-level analysis, they may lose confidence in the analytics team's competence, making future data-driven initiatives harder to sell.

- **Misleading product roadmaps**

Strategic planning based on paradoxical interpretations can set product development on an entirely wrong course, potentially taking months or years to correct.

- **Competitive disadvantage**

While a team is optimizing for the wrong segments or features based on Simpson's paradox misinterpretations, competitors with better analysis might correctly identify the real opportunities.

The Benefits of High Dimensional Analytics in E-Commerce



More accurate promotional targeting

Understanding Simpson's paradox helps e-commerce companies avoid misallocating marketing dollars. A promotion might seem effective overall but actually decreases conversion for high-value customer groups. Proper segmentation prevents targeting the wrong audiences and ensures promotional resources go where they'll generate the highest ROI.



More accurate pricing strategies

When analyzing price elasticity, companies can use Simpson's paradox to explain how what appears to be optimal pricing overall might be suboptimal for key customer groups. Companies that properly segment their analysis can implement more sophisticated pricing strategies that maximize revenue across different product categories and customer types.



Better personalization algorithms

Teams can optimize their e-commerce recommendation engines more effectively by accounting for Simpson's paradox. For example: What looks like an improvement in overall click-through rates might mask declining engagement from specific user segments. By recognizing and accounting for this, companies can build truly personalized experiences that boost sales across all customer groups.



Improved A/B testing frameworks

E-commerce platforms constantly run tests on UI changes, checkout flows, and product displays. Awareness of Simpson's paradox will help teams build testing frameworks that automatically segment results by relevant variables like device type, user history, or geographic location, preventing misleading conclusions that could negatively impact conversion rates.

How to Avoid Simpson's Paradox

- **High-dimensional analysis of customer behavior and experience**

Always analyze data by combinations of relevant dimensions like traffic source, device type, customer segments, and time periods before drawing conclusions from aggregate numbers.

- **Control for confounding variables**

Account for factors that might skew results, such as seasonality, marketing campaigns, or product mix changes.

- **Use stratified analysis**

Compare like-to-like groups rather than overall averages. For example, compare mobile conversion rates separately from desktop rather than blending them.

- **Validate with multiple cuts**

If aggregate data shows one trend, verify it holds true across major customer segments, channels, and time frames.

- **Apply statistical controls**

Use techniques like propensity score matching or regression analysis to isolate the true effect you're measuring.

The core principle: Never trust aggregate metrics alone. You need a modern analytics tool to surface deep dimensional insights across cohorts of customers.

Your competitors have a top-line number and shallow insights. You don't have to stop there.

Simpson's paradox shows that data doesn't lie—but it can mislead.

In e-commerce, where decisions are driven by metrics, the stakes are too high to rely on aggregate views alone. Use high-dimensional analytics to measure your customers' true experience. Break it down by behavior, channel, device, and state. If the story changes when you slice it, you've found something worth paying attention to.

Patented Conviva **Time-State Technology** goes beyond showing *what happened*—it **reveals the patterns, sequences, and timing across each flow**, or step, in the customer journey that led to a specific outcome. This rich context helps teams detect anomalies and reduce friction to boost conversion.

Conviva **high-dimensional analytics** continuously analyzes millions of attribute combinations in real time—pinpointing **overperforming** and **underperforming customer cohorts** that traditional averages miss to **unlock more informed growth decisions fast**.

To learn more about how Conviva can help improve the performance of your digital services, visit www.conviva.com or sign up for a demo to see what your data is really saying.



Conviva helps the world's top brands to identify and act on growth opportunities across mobile and web apps, video streaming services, and AI-driven experiences. Our real-time performance analytics platform transforms every customer interaction into actionable insight, connecting experience, engagement, and technical performance to business outcomes. By analyzing client-side session data from all users as it happens, Conviva reveals not just what happened, but how long it lasted and why it mattered—surfacing behavioral and experience patterns that give teams the context to retain more customers, resolve issues faster, and grow revenue.

To learn more about how Conviva can help improve the performance of your digital services, visit www.conviva.com, our [blog](#), and follow us on [LinkedIn](#). Curious to learn how you can identify and resolve hidden conversion issues and discover five times more opportunities for growth? Let us show you. Sign up for a [demo](#) today.