

# What advanced analytics teams are doing that you aren't

Data and analytics teams perennially face a burning — yet often unspoken — question: what drives high value actions?



Duncan Gilchrist



Jeremy Hermann



At every company I (Duncan) have worked at, the data science team faced a burning — yet often unspoken — question: **what drives high value actions?**

At Wealthfront, we obsessed over what led customers to transfer their entire external investment accounts over to us. At Uber, we dissected the factors behind frequent, long trips. And at Gopuff, we focused on understanding what drove subscriptions and basket size.

Questions like these burn because their answers are game-changing. If you can identify the critical user behaviors that drive high value actions like conversions and revenue, then you can steer the entire business to optimize for those pivotal outcomes. These insights inform product development, marketing strategies, and decision-making across the org. They can even be used to power experiences directly — e.g. as input into ranking, pricing, or fraud models — to drive users to do the things that maximize lifetime value.

In other words, insight into what drives high-value actions is the holy grail of analytics.

And yet, these questions are unspoken because they are simply so hard to answer. Think of a user's decision to upgrade to a premium tier. That could depend on their usage patterns, the features they've engaged with, their experiences with those features, the time of day, their location, the device they're using, the marketing emails they've opened, and countless other factors — all interacting in complex ways.

If you ask most data leaders, *What drives your users to take the highest value actions in your product?*, they will gaze at you with a pained look on their face. They'll probably respond, simply, *That's a hard question*. And they're right — it's an incredibly complex puzzle.

Traditional analytics tools like SQL queries or BI dashboards are great for straightforward reporting: *How many users signed up last week? What's our average revenue per user? How does revenue per user compare between iOS and Android?* But when it comes to untangling the web of factors that lead to high value actions, they're the wrong tool for the job. Answering these questions requires high-dimensional causal factor analysis, decomposing outcomes across dozens, or hundreds, or even thousands of input variables.

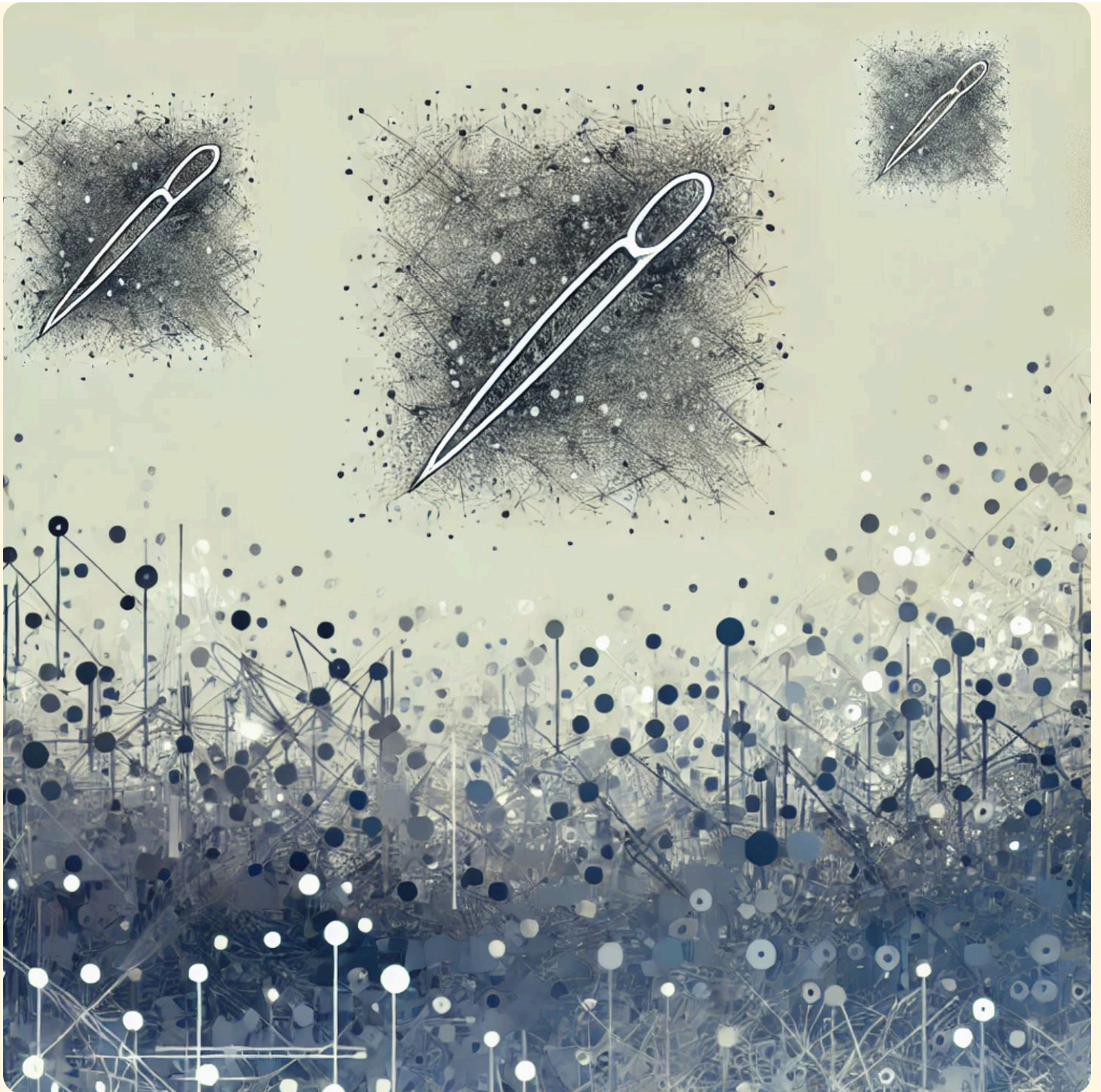
You simply can't do this in a spreadsheet or a 2x2 matrix.

## Machine learning for product analytics

There is, however, a collection of methods that are purpose built for answering these kinds of questions: machine learning.

Most people think of machine learning in the context of real-time systems like feed ranking or fraud detection, or other predictive use cases that look into the future, like lead scoring or inventory forecasting.

But at its base, machine learning is the subset of AI that allows machines to learn patterns from data without explicit programming. **And just like ML can try to peer into the future, it can also help you unpack the past.** When ML is learning patterns, it's essentially conducting multi-dimensional analytics at scale — using sophisticated statistical methods to find not just one, but all the needles in the haystack.



So you can actually use ML to dig through your historical data: to automatically detect problems that customers are facing, identify segments that are underperforming, or unveil unexpected correlations between user behaviors and outcomes.

The most advanced data science teams leverage this heavily: Amazon famously has models that quantify how usage of one product drives another, which they then use to make big decisions ranging from org-level budgets to in-app ranking, using a method known as

surrogates. And Airbnb has a Future Incremental Value framework that systematically maps short-term metrics into long-term outcomes.

Time to dive in: what can you do today to upgrade your product analytics with ML? Buckle up; this post is a bit more technical than usual.

## Three core ML techniques applied to analytics

Here are three groups of techniques you've likely heard of before, but may not have realized could be applied to this context.

### **1. Use wide classification and regression models to decompose the drivers of high-value actions**

Classification and regression models are the workhorses of ML — and they can also be the workhorses of ML-powered analytics.

Let's say you want to understand what drives customer churn in your product. You can create a predictive model that ingests hundreds of different factors, and then see which bubble up as the ones most strongly associated with churn. You might discover that 50% of churn can be explained by combinations of purchases from a specific product category, specific locations, or app versions.



- Then, by digging into the variables (and families of similar variables) that are most important in that model, you know what matters most.

Beyond conversion and churn, we've seen these techniques used to understand factors driving a wide variety of critical metrics, such as:

- Upgrades behavior in app or on your website
- Key financial metrics like weekly revenue
- Operational metrics like website or app performance

Aside: If you have training in causal methods you might be worried about causality here — and there are more advanced ways to improve the causality of models like these. But our advice is to start simple. Even if you decide to eventually build a causal model, you'll want to understand what the basic data is telling you. Our experience is also that except in certain cases where non-causal estimates can be quite misleading (e.g. pricing), the standard methods are more similar to the causal results than you might think.

## **2. Use unsupervised learning to let your data define user segments for you**

Most of the time when someone defines user segments, they do it following their own intuition and experiences. It's the easy and obvious thing to do, and we're all guilty of it.

But if you use ML-based clustering techniques, then you can let the data build its own clusters — removing the bias of the person doing the clustering. This is especially powerful when you are deliberately trying to find segments that might be non-intuitive.



For example, a streaming service might use clustering to analyze user interaction data. It could help them decide what kinds of shows to produce in the future, for segments of watchers they previously didn't know existed.

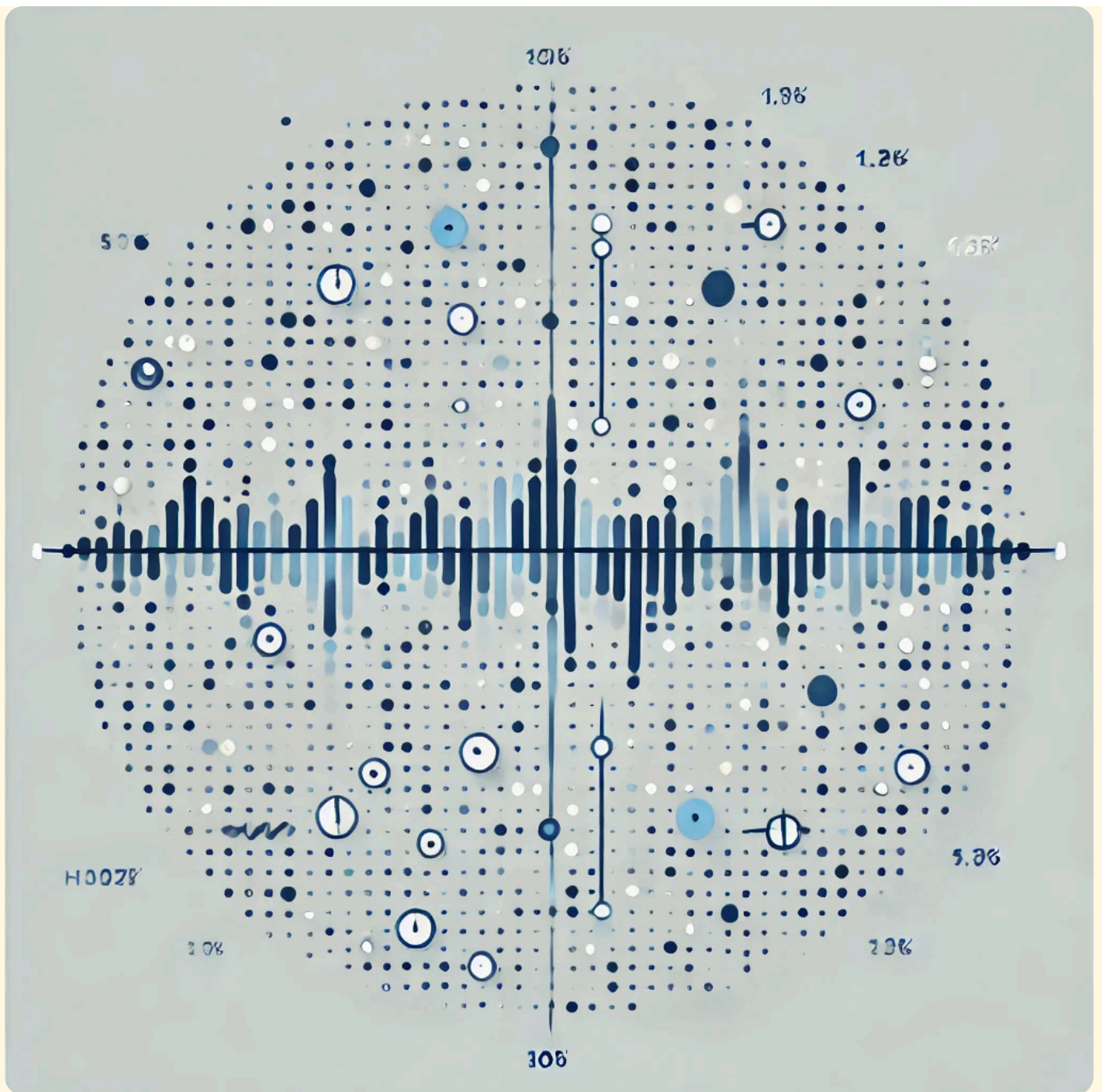
There's an exciting application of genAI here: The latest LLMs offer high dimensional embeddings, which allow you to represent text and images as high dimensional vectors -- and these too can be clustered, allowing to build clusters that embody sophisticated representations of your underlying data.

The opportunities to cluster thoughtfully are endless – you might cluster users into segments based on attributes like device type, customer lifetime value, or customer journey stage; or do a market basket analysis to understand items frequently bought together.

### **3. Use double ML to accelerate your A/B testing**

A/B testing and experimentation are crucial to product development, but are typically plagued by a common challenge: they take too long! Part of the long wait is caused by natural noise in the data — random differences between testing groups that adds variability in the numbers, so you need to wait longer until that variability averages out and you actually know if treatment beats control.

ML can actually help you squeeze out some of the noise, by controlling for key user and behavior characteristics in a principled way. As we've [written about previously](#), regularization techniques like double selection or double ML help identify the best controls to learn the right answers faster.



For instance, say you're running an A/B test for an ecommerce app that compares the effectiveness of two different promotions on click-through rates. ML allows you to control for existing factors already likely to be correlated with click-through rates, giving you a clearer picture of the true impact of each promotion.

These regularization techniques aren't limited to simple A/B tests. They can be applied to more complex experimental designs, helping you derive causal insights in scenarios where traditional randomized controlled trials aren't feasible.

# Where this is now: AI data agents do this automatically

When we wrote this post in 2024, applying these techniques to product analytics required a senior data scientist, weeks of work, and a high tolerance for the painstaking steps that come with building any model. That's changed.

AI data agents — built on frontier LLMs and grounded in an AI-managed context layer that captures what your data means and how your metrics are defined — now do this kind of multi-factor analysis directly. Ask the agent why churn went up last quarter and it doesn't just give you a number; it unpacks the drivers across user behaviors, product categories, app versions, and time periods, exactly like the wide-regression decomposition in Section 1 above. Ask about non-obvious user segments and you'll get a clustering-style breakdown without anyone explicitly running k-means.

That's what we ship at Delphina. The **Analytics Agent** handles the multi-factor breakdowns described in this post. **Deep Research** goes further — autonomously digging into questions across your data over hours, producing reports that look like what a senior analyst would deliver after a week of work. A **critic agent** keeps accuracy high by reviewing each answer before it's surfaced. And the context layer is what makes all of it reliable: institutional knowledge about metric definitions, business rules, and edge cases that the agent learns once and applies consistently.

This doesn't make the ML techniques in this post obsolete. They still matter for the model-building work data scientists do — and increasingly, AI agents themselves use these techniques under the hood. But for the very common question "what drives high value actions in our product?", a business team can now get a trustworthy answer without filing a ticket.

If you've been tackling problems like the ones discussed above, or want to see what your data looks like through an AI data agent, write us at [info@delphina.ai](mailto:info@delphina.ai) or [book a demo](#).

# Ready to unleash your data?

Discover how Delphina can transform your data science.

[Book a demo →](#)

Join our newsletter to stay up to date on features and releases.

[Subscribe](#)

By subscribing you agree to our [Privacy Policy](#) and provide consent to receive updates from our company.

## PRODUCT

[Home](#)

[Security](#)

[Integrations](#)

[Pricing](#)

## RESOURCES

[Our Story](#)

[Case Study: Substack](#)

[Case Study: BaseCamp](#)

[Customer Spotlight: Medely](#)

[Customer Spotlight: LATAM Airlines](#)

