



# Harvard Business Review

REPRINT H040LF  
PUBLISHED ON HBR.ORG  
DECEMBER 04, 2018

## ARTICLE ANALYTICS

# What Great Data Analysts Do — and Why Every Organization Needs Them

*by Cassie Kozyrkov*

ANALYTICS

# What Great Data Analysts Do — and Why Every Organization Needs Them

by Cassie Kozyrkov

DECEMBER 04, 2018



VICKI JAURON, BABYLON AND BEYOND PHOTOGRAPHY/GETTY IMAGES

The top trophy hire in data science is elusive, and it's no surprise: a “full-stack” data scientist has mastery of machine learning, statistics, and analytics. When teams can't get their hands on a three-in-one polymath, they set their sights on luring the most impressive prize among the single-origin specialists. Which of those skills gets the pedestal?

Today's fashion in data science favors flashy sophistication with a dash of sci-fi, making AI and machine learning the darlings of the job market. Alternative challengers for the alpha spot come from statistics, thanks to a century-long reputation for rigor and mathematical superiority. What about analysts?

### **Analytics as a second-class citizen**

If your primary skill is analytics (or data-mining or business intelligence), chances are that your self-confidence has taken a beating as machine learning and statistics have become prized within companies, the job market, and the media.

But what the uninitiated rarely grasp is that the three professions under the data science umbrella are completely different from one another. They may use some of the same methods and equations, but that's where the similarity ends. Far from being a lesser version of the other data science breeds, good analysts are a prerequisite for effectiveness in your data endeavors. It's dangerous to have them quit on you, but that's exactly what they'll do if you under-appreciate them.

Instead of asking an analyst to develop their statistics or machine learning skills, consider encouraging them to seek the heights of their own discipline first. In data science, excellence in one area beats mediocrity in two. So, let's examine what it means to be truly excellent in each of the data science disciplines, what value they bring, and which personality traits are required to survive each job. Doing so will help explain why analysts are valuable, and how organizations should use them.

### **Excellence in statistics: rigor**

Statisticians are specialists in coming to conclusions beyond your data safely — they are your best protection against fooling yourself in an uncertain world. To them, inferring something sloppily is a greater sin than leaving your mind a blank slate, so expect a good statistician to put the brakes on your exuberance. They care deeply about whether the methods applied are right for the problem and they agonize over which inferences are valid from the information at hand.

The result? A perspective that helps leaders make important decisions in a risk-controlled manner. In other words, they use data to minimize the chance that you'll come to an unwise conclusion.

### **Excellence in machine learning: performance**

You might be an applied machine learning/AI engineer if your response to "I bet you couldn't build a model that passes testing at 99.99999% accuracy" is "Watch me." With the coding chops to build both prototypes and production systems that work and the stubborn resilience to fail every hour for several years if that's what it takes, machine learning specialists know that they won't find the perfect solution in a textbook. Instead, they'll be engaged in a marathon of trial-and-error. Having great intuition for how long it'll take them to try each new option is a huge plus and is more valuable than an intimate knowledge of how the algorithms work (though it's nice to have both). Performance means more than clearing a metric — it also means reliable, scalable, and easy-to-maintain models that perform well in production. Engineering excellence is a must.

The result? A system that automates a tricky task well enough to pass your statistician's strict testing bar and deliver the audacious performance a business leader demanded.

### **Wide versus deep**

What the previous two roles have in common is that they both provide high-effort solutions to specific problems. If the problems they tackle aren't worth solving, you end up wasting their time and your money. A frequent lament among business leaders is, "Our data science group is useless." And the problem usually lies in an absence of analytics expertise.

Statisticians and machine learning engineers are narrow-and-deep workers — the shape of a rabbit hole, incidentally — so it's really important to point them at problems that deserve the effort. If your experts are carefully solving the wrong problems, your investment in data science will suffer low returns. To ensure that you can make good use of narrow-and-deep experts, you either need to be sure you already have the right problem or you need a wide-and-shallow approach to finding one.

### **Excellence in analytics: speed**

The best analysts are lightning-fast coders who can surf vast datasets quickly, encountering and surfacing potential insights faster than those other specialists can say "whiteboard." Their semi-sloppy coding style baffles traditional software engineers — but leaves them in the dust. Speed is their highest virtue, closely followed by the ability to identify potentially useful gems. A mastery of visual presentation of information helps, too: beautiful and effective plots allow the mind to extract information faster, which pays off in time-to-potential-insights.

The result is that the business gets a finger on its pulse and eyes on previously-unknown unknowns. This generates the inspiration that helps decision-makers select valuable quests to send statisticians and ML engineers on, saving them from mathematically-impressive excavations of useless rabbit holes.

### **Sloppy nonsense or stellar storytelling?**

"But," object the statisticians, "most of their so-called insights are nonsense." By that they mean the results of their exploration may reflect only noise. Perhaps, but there's more to the story.

Analysts are data storytellers. Their mandate is to summarize interesting facts and to use data for inspiration. In some organizations those facts and that inspiration become input for human decision-makers. But in more sophisticated data operations, data-driven inspiration gets flagged for proper statistical follow-up.

Good analysts have unwavering respect for the one golden rule of their profession: do not come to conclusions beyond the data (and prevent your audience from doing it, too). To this end, one way to spot a good analyst is that they use softened, hedging language. For example, not "we conclude" but "we are inspired to wonder". They also discourage leaders' overconfidence by emphasizing a multitude of possible interpretations for every insight.

As long as analysts stick to the facts — saying only “This is what is here.” — and don’t take themselves too seriously, the worst crime they could commit is wasting someone’s time when they run it by them.

While statistical skills are required to test hypotheses, analysts are your best bet for coming up with those hypotheses in the first place. For instance, they might say something like “It’s only a correlation, but I suspect it could be driven by ...” and then explain why they think that. This takes strong intuition about what might be going on beyond the data, and the communication skills to convey the options to the decision-maker, who typically calls the shots on which hypotheses (of many) are important enough to warrant a statistician’s effort. As analysts mature, they’ll begin to get the hang of judging what’s important in addition to what’s interesting, allowing decision-makers to step away from the middleman role.

Of the three breeds, analysts are the most likely heirs to the decision throne. Because subject matter expertise goes a long way towards helping you spot interesting patterns in your data faster, the best analysts are serious about familiarizing themselves with the domain. Failure to do so is a red flag. As their curiosity pushes them to develop a sense for the business, expect their output to shift from a jumble of false alarms to a sensibly-curated set of insights that decision-makers are more likely to care about.

### **Analytics for decision-making**

To avoid wasted time, analysts should lay out the story they’re tempted to tell and poke it from several angles with follow-up investigations to see if it holds water before bringing it to decision-makers. The decision-maker should then function as a filter between exploratory data analytics and statistical rigor. If someone with decision responsibility finds the analyst’s exploration promising for a decision they have to make, they then can sign off on a statistician spending the time to do a more rigorous analysis. (This process indicates why just telling analysts to get better at statistics misses the point in an important way. Not only are the two activities separate, but another person sits in between them, meaning it’s not necessarily any more efficient for one person to do both things.)

### **Analytics for machine learning and AI**

Machine learning specialists put a bunch of potential data inputs through algorithms, tweak the settings, and keep iterating until the right outputs are produced. While it may sound like there’s no role for analytics here, in practice a business often has far too many potential ingredients to shove into the blender all at once. One way to filter down to a promising set of inputs to try is domain expertise — ask a human with opinions about how things might work. Another way is through analytics. To use the analogy of cooking, the machine learning engineer is great at tinkering in the kitchen, but right now they’re standing in front of a huge, dark warehouse full of potential ingredients. They could either start grabbing them haphazardly and dragging them back to their kitchens, or they could send a sprinter armed with a flashlight through the warehouse first. Your analyst is the sprinter; their ability to quickly help you see and summarize what-is-here is a superpower for your process.

## **The dangers of under-appreciating analysts**

An excellent analyst is not a shoddy version of the machine learning engineer; their coding style is optimized for speed — on purpose. Nor are they a bad statistician, since they don't deal at all with uncertainty, they deal with facts. The primary job of the analyst is to say: "Here's what's in our data. It's not my job to talk about what it means, but perhaps it will inspire the decision-maker to pursue the question with a statistician."

If you overemphasize hiring and rewarding skills in machine learning and statistics, you'll lose your analysts. Who will help you figure out which problems are worth solving then? You'll be left with a group of miserable experts who keep being asked to work on useless projects or analytics tasks they didn't sign up for. Your data will lie around useless.

When in doubt, hire analysts before other roles. Appreciate them and reward them. Encourage them to grow to the heights of their chosen career (and not someone else's). Of the cast of characters mentioned in this story, the only ones that every business needs are decision-makers and analysts. The others you'll only be able to use when you know exactly what you need them for. Start with analytics and be proud of your newfound ability to open your eyes to the rich and beautiful information in front of you. Data-driven inspiration is a powerful thing.

---

**Cassie Kozyrkov** is the chief decision scientist at Google.

---