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

* [Cassie Kozyrkov](https://hbr.org/search?term=cassie%20kozyrkov)

Summary.

“Full-stack” data scientist means mastery of machine learning, statistics, and analytics. 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?Whereas excellence in statistics is about rigor and excellence in machine learning is about performance, excellence in analytics is all about speed. Analysts are your best bet for coming up with those hypotheses in the first place. 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.

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.

**Qué hacen los grandes analistas de datos y por qué todas las organizaciones los necesitan**

* [Cassie Kozyrkov](https://hbr.org/search?term=cassie%20kozyrkov)

Resumen.

Científico de datos «full-stack» significa dominio del aprendizaje automático, las estadísticas y los análisis. La moda actual en la ciencia de los datos favorece la sofisticación llamativa con un toque de ciencia ficción, lo que hace que la IA y el aprendizaje automático sean los favoritos del mercado laboral. Los retadores alternativos para el puesto alfa provienen de las estadísticas, gracias a una reputación centenaria de rigor y superioridad matemática. ¿Qué pasa con los analistas? Mientras que la excelencia en las estadísticas se basa en el rigor y la excelencia en el aprendizaje automático se basa en el rendimiento, la excelencia en el análisis se basa en la velocidad. Los analistas son su mejor opción para idear esas hipótesis en primer lugar. A medida que los analistas maduren, empezarán a acostumbrarse a juzgar lo que es importante además de lo que es interesante, lo que permitirá a los responsables de la toma de decisiones alejarse del papel de intermediario. De las tres razas, los analistas son los herederos más probables del trono de la decisión.

**Aviso:** Traducido con el uso de inteligencia artificial; puede contener errores. Participe en esta [encuesta](https://hbp.az1.qualtrics.com/jfe/form/SV_3avqQIGTlIML2Fn?source=aviso&article=https%3A%2F%2Fhbr.org%2F2018%2F12%2Fwhat-great-data-analysts-do-and-why-every-organization-needs-them) para hacernos llegar sus comentarios y obtenga información en las

La contratación de trofeos más importante en ciencia de datos es difícil de alcanzar, y no es de extrañar: un científico de datos de «pila completa» domina el aprendizaje automático, las estadísticas y el análisis. Cuando los equipos no pueden poner sus manos en un polímata tres en uno, se fijan en conseguir el premio más impresionante entre los especialistas de un solo origen. ¿Cuál de esas habilidades se queda con el pedestal?

La moda actual en la ciencia de los datos favorece la sofisticación llamativa con un toque de ciencia ficción, lo que hace que la IA y el aprendizaje automático sean los favoritos del mercado laboral. Los retadores alternativos para el puesto alfa provienen de las estadísticas, gracias a una reputación centenaria de rigor y superioridad matemática. ¿Qué pasa con los analistas?

**Analytics como ciudadano de segunda clase**

Si su habilidad principal es la analítica (o la minería de datos o la inteligencia empresarial), lo más probable es que su confianza en sí mismo se haya visto afectada a medida que el aprendizaje automático y las estadísticas se han vuelto muy apreciados en las empresas, el mercado laboral y los medios de comunicación.

Pero lo que los no iniciados rara vez comprenden es que las tres profesiones bajo el paraguas de la ciencia de datos son completamente diferentes entre sí. Pueden utilizar algunos de los mismos métodos y ecuaciones, pero ahí es donde termina la similitud. Lejos de ser una versión inferior de las otras razas de ciencia de datos, los buenos analistas son un requisito previo para la eficacia de sus esfuerzos con los datos. Es peligroso que lo abandonen, pero eso es exactamente lo que harán si los menosprecia.

En lugar de pedirle a un analista que desarrolle sus habilidades en estadística o aprendizaje automático, considere animarlo a que busque primero las alturas de su propia disciplina. En la ciencia de datos, la excelencia en un área supera a la mediocridad en dos. Así que examinemos qué significa ser realmente excelente en cada una de las disciplinas de la ciencia de datos, qué valor aportan y qué rasgos de personalidad se requieren para sobrevivir a cada trabajo. Hacerlo ayudará a explicar por qué los analistas son valiosos y cómo deben usarlos las organizaciones.

**Excelencia en las estadísticas: rigor**

Los estadísticos son especialistas en llegar a conclusiones más allá de sus datos de forma segura, son su mejor protección contra el engaño en un mundo incierto. Para ellos, inferir algo de manera descuidada es un pecado mayor que dejar la mente en blanco, así que espere que un buen estadístico frene su exuberancia. A ellos les importa mucho si los métodos aplicados son adecuados para el problema y se angustian por saber qué inferencias son válidas a partir de la información disponible.

¿El resultado? Una perspectiva que ayuda a los líderes a tomar decisiones importantes de forma controlada por el riesgo. En otras palabras, utilizan datos para minimizar la posibilidad de que llegue a una conclusión imprudente.

**Excelencia en aprendizaje automático: rendimiento**

Puede ser ingeniero de IA y aprendizaje automático aplicado si su respuesta al «Apuesto a que no podría crear un modelo que supere las pruebas con una precisión del 99,99999%» es «Mírame». Con las habilidades de codificación para crear prototipos y sistemas de producción que funcionen y la obstinada resistencia a los fallos cada hora durante varios años, si es lo que se necesita, los especialistas en aprendizaje automático saben que no encontrarán la solución perfecta en un libro de texto. En cambio, se verán enfrascados en un maratón de ensayo y error. Tener una buena intuición sobre el tiempo que les llevará probar cada nueva opción es una gran ventaja y es más valioso que un conocimiento profundo de cómo funcionan los algoritmos (aunque está bien tener ambos). El rendimiento significa más que borrar una métrica, también significa modelos fiables, escalables y fáciles de mantener que funcionan bien en producción. La excelencia en ingeniería es imprescindible.

¿El resultado? Un sistema que automatiza una tarea complicada lo suficientemente bien como para pasar las estrictas pruebas de su estadístico y ofrecer el audaz rendimiento que exigía un líder empresarial.

**Amplio contra profundo**

Lo que tienen en común las dos funciones anteriores es que ambas ofrecen soluciones de gran esfuerzo a problemas específicos. Si no vale la pena resolver los problemas que abordan, acabará haciendo perder su tiempo y su dinero. Un lamento frecuente entre los líderes empresariales es: «Nuestro grupo de ciencia de datos es inútil». Y el problema normalmente radica en la falta de experiencia en análisis.

Los estadísticos y los ingenieros de aprendizaje automático son trabajadores estrechos y profundos, con forma de madriguera de conejo, por eso es muy importante señalarles los problemas que merecen el esfuerzo. Si sus expertos resuelven cuidadosamente los problemas equivocados, su inversión en ciencia de datos sufrirá bajos rendimientos. Para asegurarse de que puede hacer un buen uso de expertos estrechos y profundos, o necesita asegurarse de que ya tiene el problema correcto o necesita un enfoque amplio y superficial para encontrarlo.

**Excelencia en la analítica: velocidad**

Los mejores analistas son codificadores ultrarrápidos que pueden navegar por vastos conjuntos de datos rápidamente, encontrar y sacar a la luz información potencial más rápido de lo que otros especialistas pueden decir «pizarra». Su estilo de codificación semidescuidado desconcierta a los ingenieros de software tradicionales, pero los deja en el polvo. La velocidad es su mayor virtud, seguida de cerca por la capacidad de identificar gemas potencialmente útiles. El dominio de la presentación visual de la información también ayuda: los gráficos bonitos y efectivos permiten a la mente extraer información más rápido, lo que vale la pena en cuanto al tiempo de obtención de información potencial.

El resultado es que la empresa toma el pulso y pone los ojos en incógnitas previamente desconocidas. Esto genera la inspiración que ayuda a los responsables de la toma de decisiones a seleccionar valiosas misiones para enviar a los estadísticos y a los ingenieros de aprendizaje automático, salvándolos de excavaciones matemáticamente impresionantes de agujeros de conejo inútiles.

**¿Tonterías descuidadas o una narración estelar?**

«Pero», objetan los estadísticos, «la mayoría de sus supuestas percepciones no tienen sentido». Con eso quieren decir que los resultados de su exploración pueden reflejar solo ruido. Quizás, pero hay más en la historia.

Los analistas son narradores de historias de datos. Su mandato es resumir datos interesantes y utilizar los datos como inspiración. En algunas organizaciones, esos hechos y esa inspiración se convierten en insumos para los responsables de la toma de decisiones humanas. Pero en las operaciones de datos más sofisticadas, la inspiración basada en los datos se marca para un seguimiento estadístico adecuado.

Los buenos analistas respetan inquebrantablemente la única regla de oro de su profesión: no llegue a conclusiones más allá de los datos (y evite que su público también lo haga). Con este fin, una forma de detectar a un buen analista es que utilicen un lenguaje suavizado y de cobertura. Por ejemplo, no «concluimos» sino «nos inspira preguntarnos». También desalientan el exceso de confianza de los líderes haciendo hincapié en una multitud de interpretaciones posibles para cada visión.

Mientras los analistas se ciñan a los hechos — diciendo solo «Esto es lo que hay aquí». — y no se tomen a sí mismos demasiado en serio, el peor crimen que podrían cometer es hacerle perder el tiempo a alguien cuando se lo digan.

Si bien se requieren conocimientos estadísticos para poner a prueba las hipótesis, en primer lugar, los analistas son su mejor opción para elaborar esas hipótesis. Por ejemplo, podrían decir algo como «Es solo una correlación, pero sospecho que podría estar impulsada por...» y luego explicar por qué piensan eso. Esto requiere una gran intuición sobre lo que podría estar sucediendo más allá de los datos y las habilidades de comunicación para transmitir las opciones a la persona que toma las decisiones, que normalmente toma las decisiones sobre qué hipótesis (de muchas) son lo suficientemente importantes como para justificar el esfuerzo de un estadístico. A medida que los analistas maduren, empezarán a acostumbrarse a juzgar lo que es importante además de lo que es interesante, lo que permitirá a los responsables de la toma de decisiones alejarse del papel de intermediario.

De las tres razas, los analistas son los herederos más probables del trono de la decisión. Dado que la experiencia en la materia contribuye en gran medida a detectar patrones interesantes en sus datos más rápido, los mejores analistas se toman en serio la idea de familiarizarse con el dominio. De lo contrario, es una señal de alerta. A medida que su curiosidad los empuja a desarrollar un sentido para el negocio, espere que su producción pase de una mezcla de falsas alarmas a un conjunto de ideas sensatamente seleccionadas que probablemente interesen a los responsables de la toma de decisiones.

**Análisis para la toma de decisiones**

Para evitar perder el tiempo, los analistas deben exponer la historia que se ven tentados a contar y tocarla desde varios ángulos con investigaciones de seguimiento para ver si se mantiene firme antes de llevarla a los responsables de la toma de decisiones. El responsable de la toma de decisiones debería funcionar entonces como un filtro entre el análisis exploratorio de datos y el rigor estadístico. Si alguien con responsabilidad de tomar decisiones considera que la exploración del analista es prometedora para una decisión que tiene que tomar, puede aprobar que un estadístico dedique tiempo a realizar un análisis más riguroso. (Este proceso indica por qué simplemente decirle a los analistas que mejoren en las estadísticas pierde el punto de una manera importante. Las dos actividades no solo son separadas, sino que otra persona se sienta entre ellas, lo que significa que no es necesariamente más eficiente que una persona haga ambas cosas).

**Análisis para el aprendizaje automático y la IA**

Los especialistas en aprendizaje automático pasan por algoritmos un montón de posibles entradas de datos, modifican la configuración y siguen iterando hasta que se producen las salidas correctas. Si bien puede parecer que aquí no hay función para el análisis, en la práctica una empresa a menudo tiene demasiados ingredientes potenciales como para meter en la licuadora de una vez. Una forma de filtrar hasta un conjunto de aportaciones prometedoras para probar es la experiencia en el dominio: pregúntele a un humano que le dé su opinión sobre cómo podrían funcionar las cosas. Otra forma es a través de la analítica. Para utilizar la analogía de la cocina, el ingeniero de aprendizaje automático es muy bueno retocando en la cocina, pero ahora mismo están frente a un enorme y oscuro almacén lleno de ingredientes potenciales. Podrían empezar a agarrarlos al azar y a arrastrarlos de vuelta a sus cocinas, o podrían enviar primero a un velocista armado con una linterna a través del almacén. Su analista es el velocista; su capacidad para ayudarlo rápidamente a ver y resumir lo que hay aquí es una superpotencia para su proceso.

**Los peligros de infravalorar a los analistas**

Un analista excelente no es una versión de mala calidad del ingeniero de aprendizaje automático; su estilo de codificación está optimizado para la velocidad, a propósito. Tampoco son malos estadísticos, ya que no se ocupan en absoluto de la incertidumbre, se ocupan de los hechos. El trabajo principal del analista es decir: «Esto es lo que hay en nuestros datos. No es mi trabajo hablar de lo que significa, pero quizás inspire a la persona que toma las decisiones a proseguir la pregunta con un estadístico».

Si hace demasiado hincapié en las habilidades de contratación y recompensas en el aprendizaje automático y las estadísticas, perderá a sus analistas. Entonces, ¿quién lo ayudará a averiguar qué problemas vale la pena resolver? Se quedará con un grupo de expertos miserables a los que se les sigue pidiendo que trabajen en proyectos inútiles o en tareas de análisis en las que no se inscribieron. Sus datos serán inútiles.

En caso de duda, contrate analistas antes que otras funciones. Apreciarlos y recompensarlos. Anímelos a crecer hasta las alturas de la carrera que han elegido (y no la de otra persona). Del elenco de personajes mencionados en esta historia, los únicos que toda empresa necesita son los responsables de la toma de decisiones y los analistas. Los demás solo los podrá utilizar cuando sepa exactamente para qué los necesita. Comience con los análisis y siéntase orgulloso de su nueva habilidad para abrir los ojos a la rica y atractiva información que tiene delante. La inspiración basada en datos es algo poderoso.