**When to Give Employees Access to Data and Analytics**

**Summary.**

As business leaders look to democratize data and analysis within their organizations, the real question they should be asking is “when” it makes the most sense. We offer the following criteria to help you determine when to empower data citizens: think about the citizen’s skill level, measure the importance of the problem, determine the problem’s complexity, empower those with domain expertise, and challenge experts to scout for bias.close

As business leaders strive to get the most out of their analytics investments, democratized data science often appears to offer the perfect solution. Using analytics software with no-code and low-code tools can put data science techniques into virtually anyone’s hands. In the best scenarios, this leads to better decision making and greater self-reliance and self-service in data analysis — particularly as demand for data scientists far outstrips their supply. Add to that reduced talent costs (with fewer high-cost data scientists) and more scalable customization to tailor analysis to a particular business need and context.

However, amid all the discussion around whether and how to democratize data science and analytics, a crucial point has been overlooked. The conversation needs to define *when*to democratize data and analytics, even to the point of redefining what democratization should mean.

Fully democratized data science and analytics presents many risks. As Reid Blackman and Tamara Sipes wrote in a [recent article](https://hbr.org/2022/12/the-risks-of-empowering-citizen-data-scientists), data science is difficult and an untrained “expert” cannot necessarily solve hard problems, even with good software. The ease of clicking a button that produces results provides no assurance that the answer is good — in fact, it could be very flawed and only a trained data scientist would know.

**It’s Only a Matter of Time**

Even with these reservations, however, democratization of data science is here to stay, as evidenced by the proliferation of [software and analytics tools](https://www.enterpriseai.news/2021/02/26/how-automl-is-democratizing-data-science-and-what-that-means-for-data-scientists/). Thomas Redman and Thomas Davenport are among those who advocate for the development of “[citizen data scientists](https://hbr.org/2021/03/4-ways-to-democratize-data-science-in-your-organization),” even screening for basic data science skills and aptitudes in every position hired.

Democratization of data science, however, should not be taken to the extreme. Analytics need not be at everyone’s fingertips for an organization to flourish. How many outrageously talented people wouldn’t be hired simply because they lack “basic data science skills?” It’s unrealistic and overly limiting.

As business leaders look to democratize data and analysis within their organizations, the real question they should be asking is “when” it makes the most sense. This starts by acknowledging that not every “citizen” in an organization is comparably skilled to be a citizen data scientist. As Nick Elprin, CEO and co-founder of Domino Data Labs, which provides data science and machine learning tools to organizations, told me in a recent conversation, “As soon as you get into modeling, more complicated statistical issues are often lurking under the surface.”

**The Challenge of Data Democratization**

Consider a grocery chain that recently used advanced predictive methods to right-size its demand planning, in an attempt to avoid having too much inventory (resulting in spoilage) or too little (resulting in lost sales). The losses due to spoilage and stockouts were not enormous, but the problem of curtailing them was very hard to solve — given all the variables of demand, seasonality, and consumer behaviors. The complexity of the problem meant that the grocery chain could not leave it to citizen data scientists to figure it out, but rather leverage a team of bona fide, well-trained, data scientists.

Data citizenry requires a “representative democracy,” as Elprin and I discussed. Just as U.S. citizens elect politicians to represent them in Congress (presumably to act in their best interests in legislative matters), so too organizations need the right representation by data scientists and analysts to weigh in on issues that others simply don’t have the expertise to address.

In short, it’s knowing when and to what degree to democratize data. I suggest the following five criteria:

**Think about the “citizen’s” skill level:** The citizen data scientist, in some shape and form, is here to stay. As stated earlier, there simply aren’t enough data scientists to go around, and using this scarce talent to address every data issue isn’t sustainable. More to the point, democratization of data is key to inculcating analytical thinking across the organization. A well-recognized example is [Coca-Cola](https://www.mckinsey.com/capabilities/operations/our-insights/coca-cola-the-people-first-story-of-a-digital-transformation), which has rolled out a digital academy to train managers and team leaders, producing graduates of the program who are credited with about 20 digital, automation, and analytics initiatives at several sites in the company’s manufacturing operations.

However, when it comes to engaging in predictive modeling and advanced data analysis that could fundamentally change a company’s operations, it’s crucial to consider the skill level of the “citizen.” A sophisticated tool in the hands of a data scientist is additive and valuable; the same tool in the hands of someone who is merely “playing around in data” can lead to errors, incorrect assumptions, questionable results, and misinterpretation of outcomes and conclusions.

**Measure the importance of the problem:**The more important a problem is to the company, the more imperative it is to have an expert handling the data analysis. For example, generating a simple graphic of historical purchasing trends can probably be accomplished by someone with a dashboard that displays data in a visually appealing form. But a strategic decision that has meaningful impact on a company’s operations requires expertise and reliable accuracy. For example, how much an insurance company should charge for a policy is so deeply foundational to the business model itself that it would be unwise to relegate this task to a non-expert.

**Determine the problem’s complexity:**Solving complex problems is beyond the capacity of the typical citizen data scientist. Consider the difference between comparing customer satisfaction scores across customer segments (simple, well-defined metrics and lower-risk) versus using deep learning to detect cancer in a patient (complex and high-risk). Such complexity cannot be left to a non-expert making cavalier decisions — and potentially the wrong decisions. When complexity and stakes are low, democratizing data makes sense.

An example is a Fortune 500 company I work with, which runs on data throughout its operations. A few years ago, I ran a [training program](https://www.forbes.com/sites/joelshapiro/2020/08/25/the-right-way-to-democratize-analytics/?sh=3aeaae5b3a6a) in which more than 4,500 managers were divided into small teams, each of which was asked to articulate an important business problem that could be solved with analytics. Teams were empowered to solve simple problems with available software tools, but most problems surfaced precisely because they were difficult to solve. Importantly, these managers were *not*charged with actually solving those difficult problems, but rather collaborating with the data science team. Notably, these 1,000 teams identified no less than 1,000 business opportunities and 1,000 ways that analytics could help the organization.

**Empower those with domain expertise:**If a company is seeking some “directional” insights — customer X is more likely to buy a product than customer Y — then democratization of data and some lower-level citizen data science will probably suffice. In fact, tackling these types of lower-level analyses can be a great way to empower those with domain expertise (i.e., being closest to the customers) with some simplified data tools. Greater precision (such as with high-stakes and complex issues) requires expertise.

The most compelling case for precision is when there are high-stakes decisions to be made based on some threshold. If an aggressive cancer treatment plan with significant side effects were to be undertaken at, for instance, greater than 30% likelihood of cancer, it would be important to differentiate between 29.9% and 30.1%. Precision matters — especially in medicine, clinical operations, technical operations, and for financial institutions that navigate markets and risk, often to capture very small margins at scale.

**Challenge experts to scout for bias:**Advanced analytics and AI can easily lead to decisions that are considered “biased.”  This is challenging in part because the point of analytics is to discriminate — that is, to base choices and decisions on certain variables. (Send this offer to this older male, but not to this younger female because we think they will exhibit different purchasing behaviors in response.) The big question, therefore, is when such discrimination is actually acceptable and even good — and when it is inherently problematic, unfair, and dangerous to a company’s reputation.

Consider the example of [Goldman Sachs](https://www.theverge.com/2021/3/23/22347127/goldman-sachs-apple-card-no-gender-discrimination), which was accused of discriminating by offering less credit on an Apple credit card to women than to men. In response, Goldman Sachs said it did not use gender in its model, only factors such as credit history and income. However, one could argue that credit history and income are correlated to gender and using those variables punishes women who tend to make less money on average and historically have had less opportunity to build credit. When using output that discriminates, decision-makers and data professionals alike need to understand how the data were generated and the interconnectedness of the data, as well as how to measure such things as differential treatment and much more. A company should never put its reputation on the line by having a citizen data scientist alone determine whether a model is biased.

Democratizing data has its merits, but it comes with challenges. Giving the keys to everyone doesn’t make them an expert, and gathering the wrong insights can be catastrophic. New software tools can allow everyone to use data, but don’t mistake that widespread access for genuine expertise.

**Cuándo dar a los empleados acceso a los datos y los análisis**

**Resumen.**Mientras los líderes empresariales buscan democratizar los datos y el análisis en sus organizaciones, la verdadera pregunta que deberían hacerse es «cuándo» tiene más sentido. Ofrecemos los siguientes criterios para ayudarlo a determinar cuándo empoderar a los ciudadanos de datos: piense en el nivel de habilidad del ciudadano, mida la importancia del problema, determine la complejidad del problema, capacite a quienes tengan experiencia en el campo y desafíe a los expertos a buscar sesgos.

Mientras los líderes empresariales se esfuerzan por aprovechar al máximo sus inversiones en análisis, la ciencia de datos democratizada suele ofrecer la solución perfecta. El uso de un software de análisis con herramientas sin código o con poco código puede poner las técnicas de ciencia de datos en manos de prácticamente cualquier persona. En el mejor de los casos, esto conduce a una mejor toma de decisiones y a una mayor autosuficiencia y autoservicio en el análisis de datos, sobre todo porque la demanda de científicos de datos supera con creces su oferta. Añada a eso la reducción de los costes de talento (con menos científicos de datos costosos) y una personalización más escalable para adaptar los análisis a una necesidad empresarial y un contexto particulares.

Sin embargo, en medio de todo el debate sobre si democratizar la ciencia de datos y el análisis y cómo hacerlo, se ha pasado por alto un punto crucial. La conversación tiene que definir cuando para democratizar los datos y la analítica, incluso hasta el punto de redefinir lo que debe significar democratización.

La ciencia y el análisis de datos totalmente democratizados presentan muchos riesgos. Como escribieron Reid Blackman y Tamara Sipes en un[artículo reciente](https://hbr.org/2022/12/the-risks-of-empowering-citizen-data-scientists), la ciencia de datos es difícil y un «experto» sin formación no puede resolver necesariamente problemas difíciles, ni siquiera con un buen software. La facilidad con la que se hace clic en un botón que produce resultados no garantiza que la respuesta sea buena; de hecho, podría tener muchos defectos y solo un científico de datos formado lo sabría.

**Solo es cuestión de tiempo**

Sin embargo, incluso con estas reservas, la democratización de la ciencia de datos llegó para quedarse, como lo demuestra la proliferación de[herramientas de software y análisis](https://www.enterpriseai.news/2021/02/26/how-automl-is-democratizing-data-science-and-what-that-means-for-data-scientists/). Thomas Redman y Thomas Davenport están entre los que abogan por el desarrollo de»[científicos de datos ciudadanos](https://hbr.org/2021/03/4-ways-to-democratize-data-science-in-your-organization)», incluso evaluando las habilidades y aptitudes básicas de ciencia de datos en cada puesto contratado.

Sin embargo, la democratización de la ciencia de datos no debe llevarse al extremo. La analítica no tiene por qué estar al alcance de todo el mundo para que una organización prospere. ¿Cuántas personas con un talento tremendo no serían contratadas simplemente porque carecen de «habilidades básicas de ciencia de datos»? No es realista y es demasiado limitante.

Mientras los líderes empresariales buscan democratizar los datos y el análisis en sus organizaciones, la verdadera pregunta que deberían hacerse es «cuándo» tiene más sentido. Esto comienza por reconocer que no todos los «ciudadanos» de una organización tienen habilidades comparables para ser científicos de datos ciudadanos. Como me dijo Nick Elprin, CEO y cofundador de Domino Data Labs, que proporciona herramientas de ciencia de datos y aprendizaje automático a las organizaciones, en una conversación reciente: «Tan pronto como se dedica a la modelización, suelen pasar desapercibidos problemas estadísticos más complicados».

**El desafío de la democratización de los datos**

Pensemos en una cadena de supermercados que recientemente utilizó métodos predictivos avanzados para dimensionar correctamente su planificación de la demanda, en un intento de evitar tener demasiado inventario (lo que provoca el deterioro) o muy poco (lo que provoca pérdidas de ventas). Las pérdidas debidas al deterioro y al agotamiento de las existencias no fueron enormes, pero el problema de reducirlas era muy difícil de resolver, dadas todas las variables de la demanda, la estacionalidad y el comportamiento de los consumidores. La complejidad del problema hizo que la cadena de supermercados no pudiera dejar que los científicos de datos ciudadanos lo descubrieran, sino más bien recurrir a un equipo de científicos de datos auténticos y bien formados.

La ciudadanía de datos requiere una «democracia representativa», como dijimos Elprin y yo. Así como los ciudadanos estadounidenses eligen a los políticos para que los representen en el Congreso (presumiblemente para que actúen en su mejor interés en los asuntos legislativos), las organizaciones también necesitan la representación correcta de los científicos y analistas de datos para opinar sobre temas que otros simplemente no tienen la experiencia para abordar.

En resumen, se trata de saber cuándo y en qué medida democratizar los datos. Sugiero los cinco criterios siguientes:

**Piense en el nivel de habilidad del «ciudadano»:** El científico de datos ciudadano, de alguna manera, llegó para quedarse. Como he dicho anteriormente, simplemente no hay suficientes científicos de datos para todos y utilizar este escaso talento para abordar todos los problemas de datos no es sostenible. Más concretamente, la democratización de los datos es clave para inculcar el pensamiento analítico en toda la organización. Un ejemplo muy reconocido es[Coca-Cola](https://www.mckinsey.com/capabilities/operations/our-insights/coca-cola-the-people-first-story-of-a-digital-transformation), que ha creado una academia digital para capacitar a los directivos y líderes de equipo, y ha producido graduados del programa a los que se les atribuyen unas 20 iniciativas digitales, de automatización y análisis en varios centros de las operaciones de fabricación de la empresa.

Sin embargo, cuando se trata de dedicarse a la modelización predictiva y al análisis avanzado de datos que podrían cambiar de manera fundamental las operaciones de una empresa, es crucial tener en cuenta el nivel de habilidad del «ciudadano». Una herramienta sofisticada en manos de un científico de datos es aditiva y valiosa; la misma herramienta en manos de alguien que se limita a «jugar con los datos» puede provocar errores, suposiciones incorrectas, resultados cuestionables y una mala interpretación de los resultados y las conclusiones.

**Mide la importancia del problema:** Cuanto más importante sea un problema para la empresa, más imperativo será contar con un experto que se encargue del análisis de datos. Por ejemplo, es probable que alguien con un panel de control que muestre los datos de una forma visualmente atractiva pueda generar un gráfico sencillo del historial de tendencias de compra. Pero una decisión estratégica que tenga un impacto significativo en las operaciones de una empresa requiere experiencia y una precisión fiable. Por ejemplo, cuánto debe cobrar una compañía de seguros por una póliza es tan fundamental para el propio modelo de negocio que no sería prudente relegar esta tarea a un no experto.

**Determine la complejidad del problema:** Resolver problemas complejos supera la capacidad del típico científico de datos ciudadano. Tenga en cuenta la diferencia entre comparar las puntuaciones de satisfacción de los clientes en todos los segmentos de clientes (métricas simples y bien definidas y de menor riesgo) y utilizar el aprendizaje profundo para detectar el cáncer en un paciente (complejo y de alto riesgo). Esa complejidad no puede dejarse en manos de un no experto que tome decisiones arrogante y, potencialmente, decisiones equivocadas. Cuando la complejidad y lo que está en juego son bajos, tiene sentido democratizar los datos.

Un ejemplo es una empresa de Fortune 500 con la que trabajo, que se basa en datos durante todas sus operaciones. Hace unos años, dirigí un[programa de entrenamiento](https://www.forbes.com/sites/joelshapiro/2020/08/25/the-right-way-to-democratize-analytics/?sh=3aeaae5b3a6a) en la que más de 4.500 directivos se dividieron en equipos pequeños, a cada uno de los cuales se le pedía que articulara un importante problema empresarial que pudiera resolverse con la analítica. Los equipos estaban capacitados para resolver problemas sencillos con las herramientas de software disponibles, pero la mayoría de los problemas surgieron precisamente porque eran difíciles de resolver. Es importante destacar que estos gerentes eran no encargado de resolver esos difíciles problemas, sino más bien de colaborar con el equipo de ciencia de datos. Cabe destacar que estos 1000 equipos identificaron no menos de 1000 oportunidades de negocio y 1000 formas en las que la analítica podría ayudar a la organización.

**Capacite a los que tienen experiencia en el campo:** Si una empresa busca información «direccional» (el cliente X tiene más probabilidades de comprar un producto que el cliente Y), entonces probablemente baste con la democratización de los datos y un poco de ciencia de datos ciudadana de nivel inferior. De hecho, abordar este tipo de análisis de nivel inferior puede ser una excelente manera de capacitar a quienes tienen experiencia en el campo (es decir, están más cerca de los clientes) con algunas herramientas de datos simplificadas. Una mayor precisión (por ejemplo, en temas complejos y de mucho en juego) requiere experiencia.

El argumento más convincente a favor de la precisión es cuando hay que tomar decisiones de alto riesgo en función de algún umbral. Si se emprendiera un plan de tratamiento agresivo del cáncer con efectos secundarios importantes con, por ejemplo, una probabilidad superior al 30% de padecer cáncer, sería importante diferenciar entre el 29,9 y el 30,1%. La precisión importa, especialmente en la medicina, las operaciones clínicas, las operaciones técnicas y para las instituciones financieras que navegan por los mercados y los riesgos, a menudo para obtener márgenes muy pequeños a gran escala.

**Desafíe a los expertos a buscar prejuicios:** La analítica avanzada y la IA pueden llevar fácilmente a decisiones que se consideran «sesgadas». Esto es difícil en parte porque el objetivo de la analítica es discriminar, es decir, basar las elecciones y las decisiones en determinadas variables. (Envíe esta oferta a este hombre mayor, pero no a una mujer más joven porque creemos que, en respuesta, mostrará diferentes comportamientos de compra). La gran pregunta, por lo tanto, es cuándo esa discriminación es realmente aceptable e incluso buena, y cuándo es intrínsecamente problemática, injusta y peligrosa para la reputación de una empresa.

Considere el ejemplo de[Goldman Sachs](https://www.theverge.com/2021/3/23/22347127/goldman-sachs-apple-card-no-gender-discrimination), que fue acusado de discriminar al ofrecer menos crédito en una tarjeta de crédito de Apple a mujeres que a hombres. En respuesta, Goldman Sachs dijo que no utilizaba el género en su modelo, solo factores como el historial crediticio y los ingresos. Sin embargo, se podría argumentar que el historial crediticio y los ingresos están correlacionados con el género y el uso de esas variables castiga a las mujeres, que tienden a ganar menos dinero de media e históricamente han tenido menos oportunidades de generar crédito. Cuando se utilizan resultados discriminatorios, tanto los responsables de la toma de decisiones como los profesionales de los datos deben entender cómo se generaron los datos y la interconexión de los datos, así como cómo medir cosas como el trato diferencial y mucho más. Una empresa nunca debe arriesgar su reputación haciendo que un científico de datos ciudadano determine por sí solo si un modelo está sesgado.

Democratizar los datos tiene sus ventajas, pero conlleva desafíos. Dar las llaves a todo el mundo no los convierte en expertos, y recopilar información equivocada puede ser catastrófico. Las nuevas herramientas de software pueden permitir a todo el mundo utilizar los datos, pero no confunda ese acceso generalizado con una experiencia genuina.