

Presentado por



Google Cloud

Informe Accelerate State of DevOps 2023



Patrocinadores Premiere



Deloitte.

Qarik

Índice

Introducción

Resumen ejecutivo	03
Conceptos y medidas	06

Capítulo 1

¿Cómo se compara?	10
-------------------------	----

Capítulo 2

Enfoque en los usuarios	17
-------------------------------	----

Capítulo 3

Las competencias técnicas conducen a un mayor rendimiento	20
---	----

Capítulo 4

La documentación es fundamental	27
---------------------------------------	----

Capítulo 5

La fiabilidad promueve el rendimiento	31
---	----

Capítulo 6

Una infraestructura flexible es crucial para el éxito	38
---	----

Capítulo 7

Nada de esto es posible si no se invierte en la cultura	45
---	----

Capítulo 8

Cómo, cuándo y por qué importa ser quien eres	51
---	----

Epílogo

Reflexiones finales	57
Agradecimientos	58
Autores	59
Metodología	62
Datos demográficos y firmográficos	72
Los modelos	81
Más información	91
Apéndice	92

Todas las citas obtenidas el 27 de septiembre del 2023

Resumen ejecutivo

El programa de investigación de DORA lleva casi una década estudiando las competencias y métricas de empresas tecnológicas con un alto rendimiento. En él han participado más de 36.000 profesionales de empresas de diferentes tamaños y sectores. Gracias por vuestras valiosas aportaciones.

El objetivo de DORA es determinar la relación que hay entre las formas de trabajar (es decir, las competencias) y los resultados obtenidos: logros significativos que sean relevantes para una empresa y para las personas que trabajan en ella. Esta investigación utiliza una rigurosa evaluación estadística y se aplica a todo tipo de plataforma (consulta [Metodología](#)).

Confiamos en que los responsables y profesionales aprovechen esta valiosa información para hacerse una idea de dónde pueden generar un impacto.

Este año, la investigación ha abordado tres resultados clave y las competencias que contribuyen a alcanzarlos:

- **Rendimiento empresarial:** la empresa no solo debe generar ingresos, sino también aportar valor a sus clientes, así como a la comunidad en general.
- **Rendimiento del equipo:** la capacidad de un equipo de aplicaciones o servicios para crear valor, innovar y colaborar.
- **Bienestar de los empleados:** las estrategias que adopta una empresa o un equipo deben beneficiar a los empleados y contribuir a reducir el desgaste, fomentar una experiencia laboral satisfactoria y aumentar la capacidad que tienen las personas de producir resultados valiosos (es decir, la productividad).

En la investigación también se han analizado los medios o medidas de rendimiento habituales como un fin en sí mismos:

- **Rendimiento del envío de software:** los equipos pueden cambiar sus sistemas tecnológicos de forma rápida, segura y eficiente.
- **Rendimiento operativo:** el servicio ofrece una experiencia fiable a sus usuarios.



Principales conclusiones



Establecer una cultura sana

La cultura es fundamental para desarrollar competencias técnicas, estimular el rendimiento técnico, alcanzar los objetivos de rendimiento de la empresa y ayudar a los empleados a tener éxito. Los equipos con culturas generativas tienen un rendimiento empresarial un 30 % mayor.



Crear sin perder de vista al usuario

Un enfoque centrado en el usuario sirve para detectar e impulsar mejoras en todas las competencias técnicas, en procesos y culturales que estudiamos en nuestra investigación. Los equipos se pueden desplegar todo lo rápido y bien que quieran, pero, si no se tiene en cuenta al usuario, es posible que no sirva de nada. Los equipos que se centran en el usuario tienen un rendimiento empresarial un 40 % mayor.



Activar el rendimiento del envío de software con revisiones de código más rápidas

Una de las vías más efectivas para mejorar el rendimiento del envío de software es acelerando las revisiones de código. Los equipos que revisan código más rápido tienen un rendimiento del envío de software un 50 % mayor.



Amplifica las competencias técnicas gracias a una documentación de calidad

La documentación de calidad amplifica el efecto de las competencias técnicas en el rendimiento de la empresa. Por ejemplo, se calcula que el desarrollo basado en troncales tiene un impacto 12,8 veces mayor en el rendimiento empresarial si se cuenta con documentación de calidad frente a cuando la documentación es de baja calidad.



Aumentar la flexibilidad de la infraestructura con la nube

El cloud computing es beneficioso porque crea una infraestructura flexible. Por ejemplo, usar una nube pública conlleva un aumento del 22 % en la flexibilidad de la infraestructura. Esa flexibilidad, a su vez, contribuye a que el rendimiento de la empresa sea un 30 % mayor que el de las infraestructuras inflexibles. Si queremos sacar el máximo partido a la nube, hay que aprovechar las características y competencias diferenciadoras que ofrece, en concreto, la flexibilidad de la infraestructura.



Buscar el equilibrio entre la velocidad de envío, el rendimiento operativo y el enfoque en el usuario

Para que el rendimiento empresarial alcance su máximo potencial, se necesita tanto un buen rendimiento del envío de software como un buen rendimiento operativo. Mantener el equilibrio entre ambos y un enfoque centrado en el usuario permitirá obtener excelentes resultados organizativos y, al mismo tiempo, mejorará el bienestar de los empleados.



Distribuir el trabajo de manera justa

Las personas que se identifican como pertenecientes a un grupo infrarrepresentado, las mujeres y las personas que autodeterminan su género tienen unos niveles más altos de desgaste. Esto posiblemente se deba a numerosos factores del sistema y del entorno. Como era de esperar, observamos que los encuestados que realizan más trabajo repetitivo son los que más posibilidades tienen de sufrir un mayor desgaste, y que los miembros de grupos infrarrepresentados tienen más posibilidades de hacer más trabajo repetitivo. Los encuestados que pertenecen a grupos infrarrepresentados tienen un desgaste un 24 % mayor que los que no están infrarrepresentados. Además, los encuestados de grupos infrarrepresentados realizan un 29 % más de trabajo repetitivo. Las mujeres o las personas que autodeterminan su género realizan un 40 % más de trabajo repetitivo que los hombres.

Aplicar las estadísticas de DORA en tu contexto

Los equipos que adoptan una mentalidad y una práctica de mejora continua suelen ser los que más se benefician.¹ DORA te puede ayudar a influir en tus iniciativas de mejora.

Para sacar el máximo partido a esta investigación, aplícalas al contexto de tu propio equipo y a tus usuarios. Por ejemplo, antes hemos mencionado que los equipos que revisan el código más rápido tienen un rendimiento del envío de software un 50 % mayor. Sin embargo, es poco probable que el rendimiento del envío de software mejore si tus revisiones de código son rápidas pero la velocidad está limitada en otras partes del sistema. La investigación se puede contextualizar cuando los profesionales hablan sobre cómo se trabaja hoy en día. Estas conversaciones pueden conducir a una mejora de la empatía, la colaboración y la comprensión de las motivaciones de todos los participantes.

El trabajo de mejora nunca se termina. Identificas un cuello de botella en tu sistema, lo gestionas y repites el proceso. Las comparaciones más significativas se hacen observando la misma aplicación en diferentes momentos, no fijándote en otras aplicaciones, empresas o sectores.

Métricas y mediciones

Gracias a las métricas y los paneles, los equipos pueden monitorizar su progreso y el camino correcto.

Los profesionales y líderes buscan rendimiento empresarial, rendimiento del equipo y bienestar. Pero el objetivo no es medirlos, como tampoco lo es enviar software.

Obsesionarse con las métricas de rendimiento puede llevar a comportamientos inefficientes. Es mejor invertir en competencias y aprendizaje para alcanzar el éxito. Los equipos que aprenden más también mejoran más.

No puedes mejorar en solitario

Aprendemos de la experiencia de los demás. El sitio web de DORA Community es un excelente foro para compartir y conocer iniciativas de mejora: <https://dora.community>.

¹Informe Accelerate State of DevOps 2022. <https://dora.dev/research/2022/dora-report/2022-dora-accelerate-state-of-devops-report-es.pdf#page=7>



Conceptos y medidas

En esta sección se describen los conceptos que DORA intenta medir.¹ Tanto el informe como nuestros modelos se basan en estos elementos. Aquí se incluyen los ingredientes que hemos utilizado para elaborar este informe. Por ello, como autores, es importante que tengamos claros estos conceptos, así como mantener la coherencia al hablar de ellos. Las tablas que aparecen a continuación pretenden aclarar ideas y sentar las bases para todos, lectores y autores por igual, con una terminología compartida.

Muchos de los conceptos de este informe son polifacéticos, por lo que solemos usar diversos indicadores para reflejarlos correctamente. Podemos evaluar si hemos reflejado correctamente estos conceptos con análisis de factores exploratorios y confirmatorios. Si quieras saber más sobre estos procesos, consulta el apartado [Metodología](#). Una vez evaluados nuestros métodos de medición, escalamos las puntuaciones de 0 a 10, donde 0 representa la total falta de presencia de un concepto y 10 su máxima presencia. Nuestra intención era estandarizar la forma de referirnos al funcionamiento de estos conceptos y facilitar la comparación de datos a lo largo de los años.

Cada concepto del que hablemos irá acompañado de la siguiente información:

- Un ícono que ayude a transmitir el significado y facilite su localización cuando se use este capítulo como referencia.
- La puntuación promedio de ese concepto en la muestra (la media).
- Los límites del rango intercuartil (IQR). Al proporcionar los dos números (percentiles 25 y 75) en los que se ubica el 50 % central de los datos, estos límites deberían servir para transmitir la dispersión de las respuestas.
- El valor del medio es un conjunto de datos (la mediana). Si el valor difiere enormemente de la media, puede indicar que los datos son sesgados.
- Una descripción del concepto y cómo lo medimos.

¹Las preguntas de la encuesta utilizadas en nuestro análisis están publicadas en <https://dora.dev>

Resultados clave

Los resultados clave son los objetivos que consideramos que las personas, los equipos o las empresas se esfuerzan por alcanzar (por ejemplo, el rendimiento empresarial) o evitar (por ejemplo, el desgaste). Por ello, creemos que las medidas son un buen recurso para que las personas se evalúen a sí mismas, a sus equipos y a sus empresas.

Rendimiento empresarial	Rendimiento del equipo	Rendimiento del envío de software
Media 6,3	IQR 5-8	Media 6,3
Mediana 6,3	Mediana 8	Mediana 6,4
Las empresas con un alto rendimiento tienen más clientes, mayores beneficios y una cuota de mercado relativa mayor para su producto o servicio principal.	Los equipos con un alto rendimiento se adaptan al cambio, se apoyan unos en otros, trabajan de manera eficiente, innovan y colaboran.	Las siguientes cuatro métricas miden la velocidad y estabilidad del envío de software:
Rendimiento operativo	Objetivos de fiabilidad	<ul style="list-style-type: none">• Frecuencia de despliegue• Plazo de entrega de los cambios• Índice de fallos por cambios• Tiempo de recuperación tras un despliegue fallido
Media 6,2	IQR 5-7,5	Media 7
Mediana 6,3	Mediana 7,5	
El grado en el que un servicio puede cumplir las expectativas de sus usuarios, lo que incluye medidas como la disponibilidad y el rendimiento.	El grado en el que un servicio cumple sus objetivos declarados en relación con medidas como la disponibilidad, el rendimiento y la exactitud.	

El **bienestar** es una combinación de desgaste, productividad y satisfacción laboral

Desgaste	Productividad	Satisfacción laboral
Media 4,1	IQR 2-6	Media 6,08
Mediana 4	Mediana 7,5	Mediana 5,7-7,1
No se trata solo del peaje psicológico y físico laboral, sino de cómo cada persona percibe el valor y la relevancia de su trabajo. El desgaste desencadena escepticismo. ²	Una persona productiva realiza un trabajo que se ajusta a sus habilidades, crea valor y le permite trabajar de manera eficiente.	Una pregunta de ítem único que pide al encuestado que tenga todo en cuenta para valorar cómo se siente con respecto a su trabajo en general. ³

² Maslach C, Leiter MP. Understanding the burnout experience: recent research and its implications for psychiatry. World Psychiatry. 2016 Jun;15(2):103-11. doi: 10.1002/wps.20311. PMID: 27265691; PMCID: PMC4911781.

³ Warr, P., Cook, J. y Wall, T. "Scales for the measurement of some work attitudes and aspects of psychological well-being". Journal of Occupational Psychology, 52(2), 1979. 129-148. <https://doi.org/10.1111/j.2044-8325.1979.tb00448.x>

Procesos y competencias técnicas

Son actividades, prácticas o estados que pueden surgir en un equipo o empresa.

Es decir, lo que hacen los equipos o cómo son.

 Contribución de la inteligencia artificial Media IQR Mediana 3,3 0,3-6,3 2,4 La importancia del papel de la inteligencia artificial como ayuda en distintas tareas técnicas.	 Arquitectura con bajo acoplamiento Media IQR Mediana 6,4 4,7-8,3 6,7 Software que se puede escribir, probar y desplegar de forma independiente.	 Integración continua Media IQR Mediana 6,9 5-8,9 7,8 La práctica que consiste en desarrollar y probar automáticamente cambios de software.
 Entrega continua Media IQR Mediana 7,0 5,7-8,7 7,3 La capacidad para incluir cambios de todo tipo, como nuevas funciones, cambios en la configuración, correcciones de errores y experimentos, en la fase de producción o en manos de los usuarios de una forma no solo rápida y segura, sino también sostenible. ⁴	 Velocidad de revisión de código Media IQR Mediana 6,5 6-8 6 Un único elemento que evalúa cuánto tiempo se tarda desde la solicitud de extracción hasta la revisión de cambio de código.	 Documentación Media IQR Mediana 5,8 3,8-7,9 6,25 La calidad del contenido escrito que el personal de la empresa crea y usa en su trabajo diario.
 Prácticas de fiabilidad Media IQR Mediana 5,9 3,9-8,3 6,1 Actividades y prácticas que utilizan los equipos para mejorar el rendimiento operativo de los servicios.	 Desarrollo basado en troncales Media IQR Mediana 5,6 3,9-7,8 5,6 La práctica que consiste en hacer cambios pequeños y frecuentes que se fusionan periódicamente en la rama principal del código del sistema de control de versiones.	 Infraestructura flexible Media IQR Mediana 6,6 5-8,3 7,3 Infraestructura escalable que es elástica, accesible y medida. ⁵

⁴"What is Continuous Delivery" <https://continuousdelivery.com/>

⁵Instituto Nacional de Normas y Tecnología. (2018) NIST. The NIST Definition of Cloud Computing. Publicación disponible en <https://csrc.nist.gov/pubs/sp/800/145/final>

Aspectos relativos a la cultura

Definir el concepto de cultura no es fácil, pero podríamos decir que esta incluye las normas imperantes (como la flexibilidad), la orientación predominante (como el enfoque centrado en el usuario) y el ambiente (como la estabilidad de la empresa) del lugar de trabajo.

 Distribución del trabajo	 Flexibilidad	 Seguridad laboral
Media 5,8	IQR 3,8-7,9	Mediana 5,8
Procesos formales para ayudar a los empleados a distribuir las tareas de manera equitativa dentro de un equipo.		Cómo, dónde y cuándo trabaja una persona en sus tareas. ⁶
Media 5,9	IQR 3,3-8,3	Mediana 6,7
Una métrica de ítem único que pregunta a las personas con cuánta frecuencia se preocupan por su seguridad laboral. Las puntuaciones más altas indican una preocupación menor.		
 Estabilidad empresarial	 Intercambio de conocimientos	 Enfoque centrado en el usuario
Media 7,2	IQR 6,7-8,3	Mediana 8,3
Una métrica de ítem único que pregunta lo estable o inestable que es el entorno laboral para los empleados.		La forma de difundir ideas e información en una empresa. Los miembros del equipo responden a las preguntas una vez y ponen la información a disposición de los demás. No hay que esperar para obtener una respuesta. ⁷
Media 7,8	IQR 5,6-8,3	Mediana 7,8
Entender e incorporar las necesidades y los objetivos de los usuarios para mejorar los productos y servicios. ⁸		
 Cultura empresarial de Westrum		
Media 7,3	IQR 6,1-8,6	Mediana 7,8
La forma habitual que tienen las empresas de responder a problemas y oportunidades. Hay tres tipos de cultura: generativa, burocrática y patológica. ⁹		

⁶Shifrin, Nicole V. y Jesse S. Michel. "Flexible work arrangements and employee health: A meta-analytic review". Work & Stress 36, n.º 1, 2022. 60-85

⁷"2022 Developer Survey" <https://survey.stackoverflow.co/2022#overview>

⁸Kersten, Mik. Project to Product: How to survive and thrive in the age of digital disruption with the flow framework (IT Revolution, 2018), 54. <https://itrevolution.com/product/project-to-product/>

⁹Westrum R. "A typology of organisational cultures". BMJ Quality & Safety, 2004. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1765804/>

¿Cómo se compara?

Ley de Goodhart: cuando una medida se convierte en un objetivo, deja de ser una buena medida.¹

Datos clave

Lo primero que hay que hacer para mejorar el rendimiento es definir un valor de referencia para el rendimiento del envío de software, el rendimiento operativo y el enfoque centrado en el usuario de una aplicación en ese momento. Estas medidas ayudan a los equipos a evaluar cómo lo están haciendo y son un buen indicador de los cambios que se van produciendo a lo largo de tiempo.

No obstante, estas medidas no son los medios por los que un equipo mejora. Con este valor de referencia, es importante evaluar la fortaleza del equipo en el contexto de una amplia variedad de personas, procesos y competencias técnicas para identificar qué es lo que podría estar impidiendo que se avance.² Por otro lado, los equipos necesitan tiempo y espacio para unificar criterios, experimentar y reevaluar su situación. Repetir este proceso ayudará a los equipos a adoptar una mentalidad y una práctica de mejora continua.

¹Strathern, Marilyn (1997). "Improving ratings": audit in the British University system". European Review. John Wiley & Sons. 5 (3): 305–321. doi:10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4. S2CID 145644958.

²Este informe y los recursos que aparecen en "Capability catalog" (<https://dora.dev/devops-devops-capabilities/>) pueden resultar útiles.

³Forsgren, N., Storey, M-A. et. al. "The SPACE of Developer Productivity: There's more to it than you think", 2021. <https://queue.acm.org/detail.cfm?id=3454124>

Ten cuidado con estos y otros errores en comparaciones como estas:

- **Comparaciones desiguales:** es probable que no sirva de nada comparar aplicaciones basándose únicamente en estos grupos. Al hacerlo, se descarta el contexto de cada aplicación, lo que podría perjudicar el objetivo de mejorar.
- **Establecer métricas como objetivos:** ignorar la ley de Goodhart y hacer afirmaciones generales como que "todas las aplicaciones deben demostrar un rendimiento de élite a final de año" aumenta las posibilidades de que los equipos intenten manipular las métricas.
- **Una métrica para todo:** tratar de medir sistemas complejos con la "única métrica importante". Debemos usar una combinación de métricas para mejorar la comprensión.³
- **Métricas de alcance limitado:** es habitual medir lo que es más fácil de medir y no lo más significativo.
- **Usar el sector como un escudo contra la mejora:** por ejemplo, puede que algunos equipos de sectores muy regulados utilicen la normativa como excusa para no alterar su statu quo.

Para obtener más información sobre nuestras conclusiones y consejos para impulsar una mentalidad y una práctica de mejora continua, consulta "How to transform" en: dora.dev/devops-capabilities/cultural/devops-culture-transform/

Introducción

Todos los años realizamos uno o varios análisis de grupos para detectar las tendencias más frecuentes en las aplicaciones. Te recomendamos que los uses para entender cómo lo estáis haciendo, aunque no debes obsesionarte con las comparaciones. Las mejores comparaciones son las que se hacen pasado un tiempo en las mismas aplicaciones, no en otras diferentes, cuyo contexto siempre será distinto.

Los equipos desarrollan software para los usuarios, y ellos son los que juzgarán la fiabilidad y utilidad del servicio en última instancia. Los equipos que se centran en las necesidades de los usuarios están mejor preparados para desarrollar una solución acertada. Al combinar el enfoque en el usuario con el rendimiento del envío de software y el rendimiento operativo, esos equipos también están preparados para desarrollar bien esa solución.

Los equipos que se centran en las necesidades de los usuarios desarrollan lo correcto y lo hacen bien.

Resultados

Rendimiento del envío de software

Utilizamos las siguientes métricas para evaluar el rendimiento del envío de software:

- **Plazo de entrega de los cambios:** cuánto tarda un cambio en pasar de confirmado a desplegado.
- **Frecuencia de despliegue:** con qué frecuencia los cambios se envían a producción.
- **Índice de fallos por cambios:** con qué frecuencia un despliegue introduce un error que requiere una intervención inmediata.
- **Tiempo de recuperación tras un despliegue fallido:** cuánto tarda la recuperación de un despliegue fallido.

Un método habitual para mejorar las cuatro métricas es disminuir el tamaño del lote de cambios de una aplicación.⁴ Cuanto más pequeños son los cambios, más fácil es trabajar con ellos y pasar al proceso de entrega. Además, la recuperación es más fácil si se produce un fallo. Los equipos deberían hacer cambios lo más pequeños posible para que el proceso de entrega sea rápido y estable. Esta forma de trabajar contribuye tanto a la velocidad como a la estabilidad de los cambios.

Este año hemos ajustado las métricas de rendimiento del envío de software. Si quieres saber más sobre estos

⁴A menudo, una función se puede desglosar en numerosos cambios que se ofrecen de forma independiente. Nuestras métricas de rendimiento del envío de software evalúan los cambios realizados en una aplicación o servicio.

cambios, consulta el apartado "Redefinir cómo medimos el rendimiento del envío de software" del [Apéndice](#).

A continuación puedes comprobar cómo les va a los encuestados de este año con el rendimiento del envío de software:

Nivel de rendimiento	Frecuencia de despliegue	Plazo de entrega de los cambios	Índice de fallos por cambios	Tiempo de recuperación tras un despliegue fallido	Porcentaje de encuestados
Élite	Bajo demanda	Menos de un día	5 %	Menos de una hora	18 %
Alto	Entre una vez al día y una vez a la semana	Entre un día y una semana	10 %	Menos de un día	31 %
Intermedio	Entre una vez a la semana y una vez al mes	Entre una semana y un mes	15 %	Entre un día y una semana	33 %
Bajo	Entre una vez a la semana y una vez al mes	Entre una semana y un mes	64 %	Entre uno y seis meses	17 %



Rendimiento operativo

Hemos evaluado el rendimiento operativo preguntando a los encuestados la frecuencia de las siguientes afirmaciones sobre el servicio:

- Recibe informes acerca de lo insatisfechos que están los usuarios finales con la fiabilidad del sistema.
- No está disponible, su rendimiento es inferior a lo esperado o no funciona correctamente.

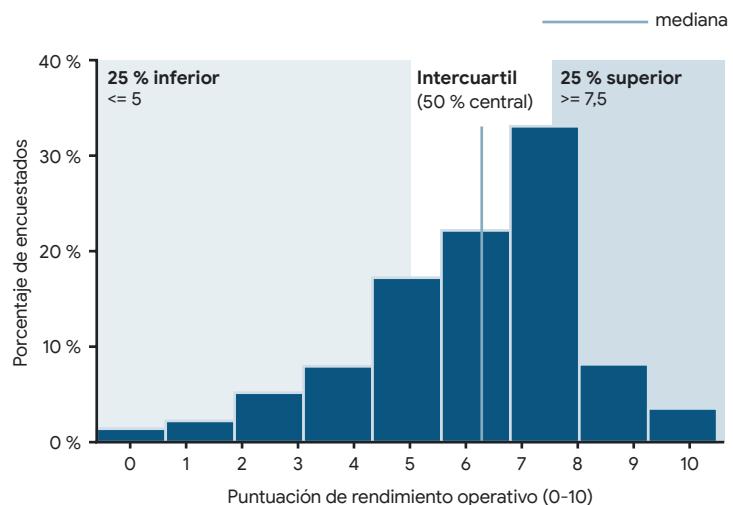
Para descubrir cómo el rendimiento operativo influye positivamente en el rendimiento empresarial, consulta el [Capítulo 5 - La fiabilidad promueve el rendimiento](#).

Enfoque centrado en el usuario

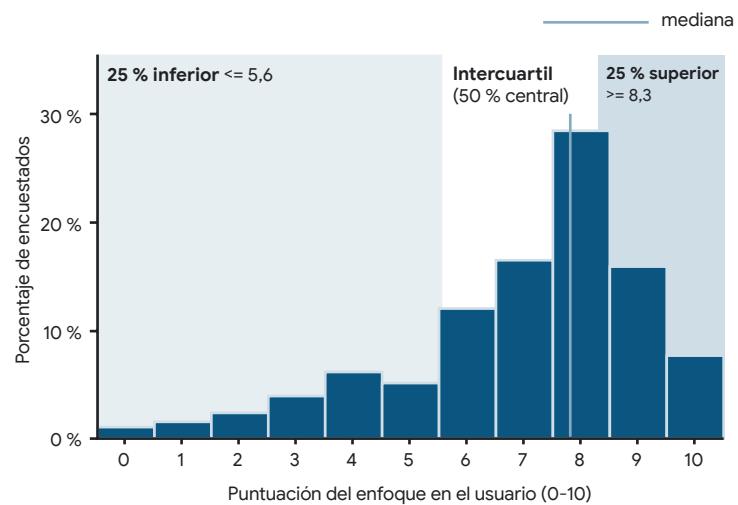
Las aplicaciones o los servicios centrados en los usuarios se desarrollan pensando en el usuario final. Para desarrollar un producto así, hay que identificar bien lo que los usuarios necesitan y saber incorporarlo en la hoja de ruta del producto. Hemos evaluado el enfoque centrado en los usuarios de los encuestados preguntándoles hasta qué punto eran verdad las siguientes afirmaciones:

- El equipo entiende a la perfección lo que los usuarios quieren conseguir.
- El éxito del equipo se evalúa según el valor que aporte a la empresa y a los usuarios de la aplicación.
- Las especificaciones (como la planificación de requisitos) se revisan constantemente y priorizan según las señales de los usuarios.

A continuación puedes comprobar cómo les va a los encuestados de este año con el rendimiento operativo:



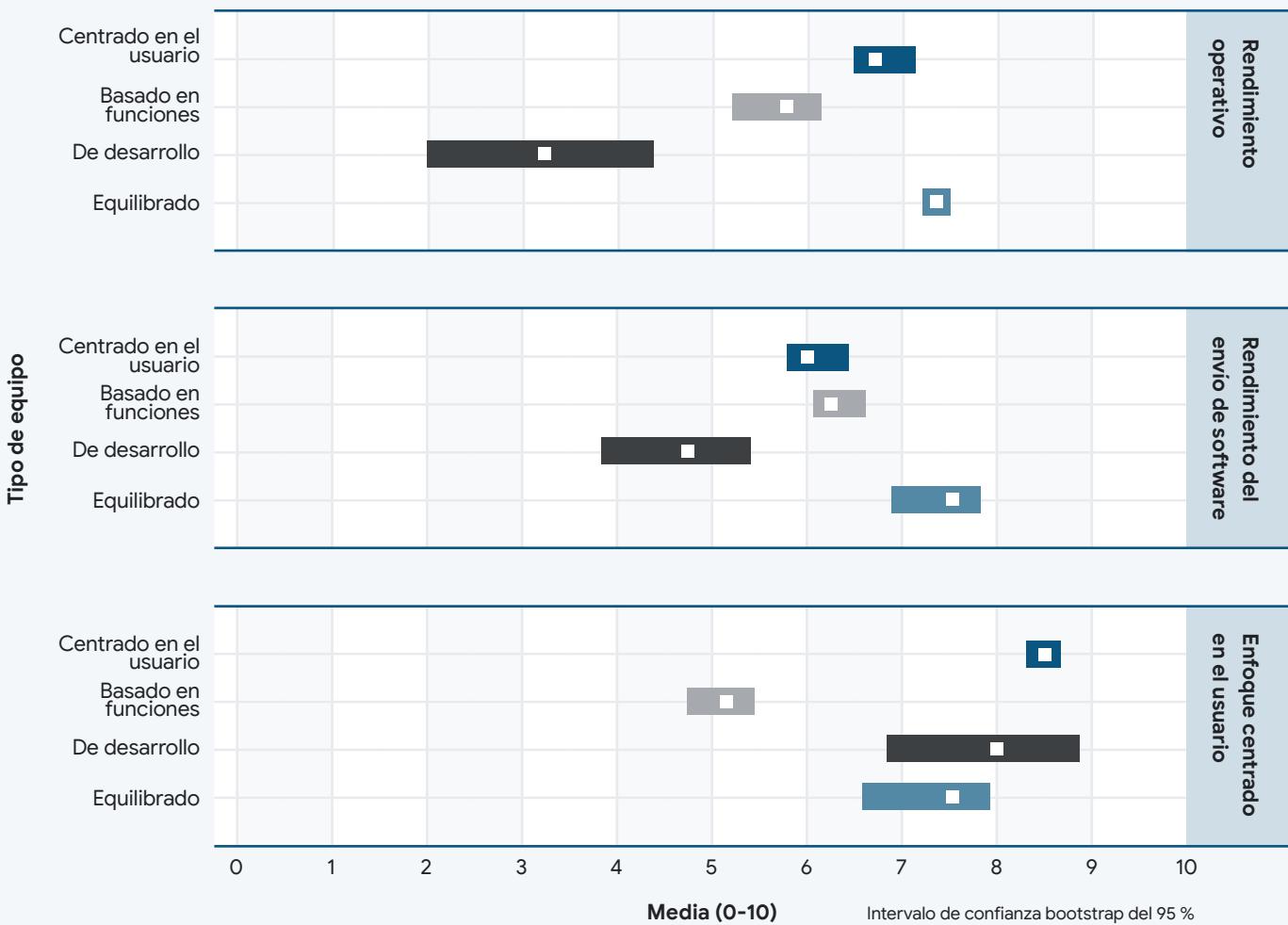
Y aquí puedes ver cómo les va con el enfoque centrado en los usuarios:



Agrupación en clúster por tipo de equipo

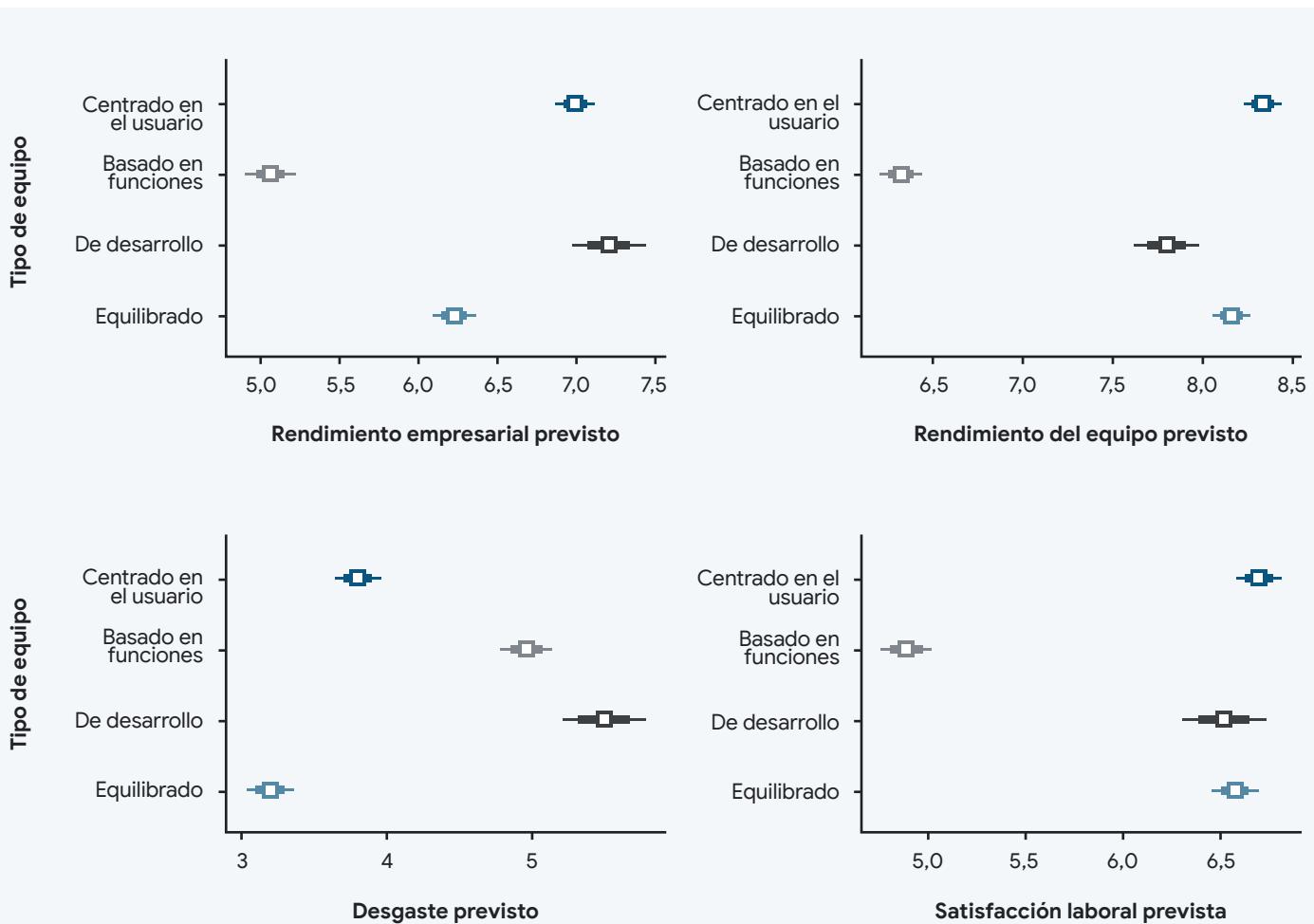
La comparación del rendimiento del envío de software, el rendimiento operativo y el enfoque centrado en el usuario como unidad da lugar a cuatro tipos de equipos que, al igual que todas las medidas implicadas en su creación, se refieren a una aplicación o un servicio.

Hemos asignado estos nombres a los tipos de equipos: Centrado en el usuario, Basado en funciones, De desarrollo y Equilibrado.



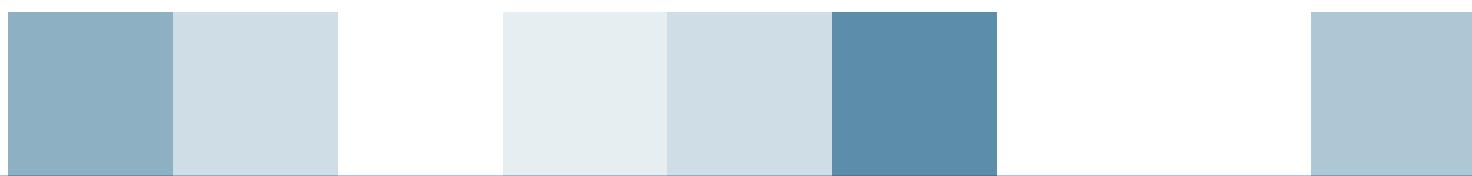
Las métricas de rendimiento de las que hemos estado hablando son como ajustes a disposición de la empresa o del equipo para cambiar el rendimiento empresarial, el rendimiento del equipo y el bienestar de sus miembros.

Los siguientes gráficos muestran los resultados de rendimiento previstos para cada tipo de equipo.



* El punto representa el lugar estimado de la media según el tipo de equipo. El intervalo grueso es donde se sitúa el 66 % de las simulaciones. El intervalo fino es donde se sitúa el 89 % de las simulaciones.

Cada tipo de equipo tiene unas características singulares, conforma un porcentaje significativo de nuestros encuestados y obtiene resultados diferentes. Es posible que tu propio equipo no encaje al 100 % en uno solo, ni tampoco esperamos que se mantenga constante a lo largo del tiempo.



¿Qué indican estos resultados?

» **Tipo de equipo centrado en el usuario**

Este tipo de equipo es el que más se centra en las necesidades de los usuarios. Este enfoque, acompañado de un sólido rendimiento del envío de software y un sólido rendimiento operativo, augura los niveles más altos de rendimiento empresarial. Sin embargo, sí que muestra algo más de desgaste que el equipo equilibrado. Mejorar el rendimiento del envío de software o el rendimiento operativo puede ser la forma más acertada de reducir el desgaste de estos equipos.

» **Tipo de equipo basado en funciones**

Este equipo da prioridad a las funciones de envío. Centrarse de manera incesante en los envíos podría suponer una distracción para que el equipo satisfaga las necesidades de los usuarios, como así demuestran unos valores más bajos de orientación al cliente y rendimiento operativo. Este tipo de equipos registra algunos de los niveles más altos de desgaste y más bajos de satisfacción laboral, rendimiento del equipo y rendimiento empresarial. Los empleados quieren ofrecer valor, no solo funciones. Los equipos basados en funciones podrían aprovechar y tener en

cuenta las necesidades de sus usuarios como una manera de obtener más valor de las funciones que se envían.

» **Tipo de equipo de desarrollo**

Este tipo de equipo alcanza un buen rendimiento empresarial al centrarse en las necesidades de los usuarios de la aplicación. Sin embargo, es posible que aún estén desarrollando la forma de adaptar su producto al mercado o sus competencias técnicas. Suelen encontrarse más a menudo en empresas más pequeñas. Su rendimiento del envío de software y su rendimiento operativo son más bajos, y los equipos que trabajan en estas aplicaciones registran unos niveles de desgaste más altos que los trabajadores de equipos equilibrados o centrados en el usuario. Es probable que estos equipos tengan procesos tediosos o tareas laboriosas que se podrían automatizar para mejorar el rendimiento del envío de software y el rendimiento operativo.

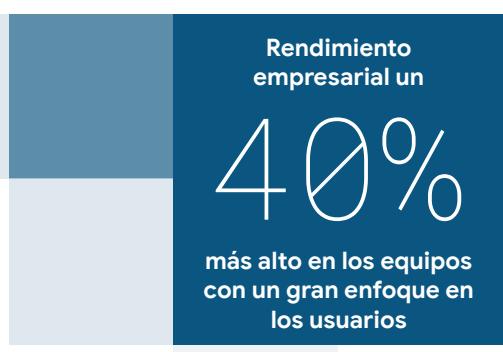
» **Tipo de equipo equilibrado**

La estrategia de este equipo es equilibrada y continuada. Utiliza la tecnología de una forma sostenible para alcanzar un buen rendimiento empresarial, un buen rendimiento del equipo y un buen nivel de satisfacción laboral. Estos equipos también registran los niveles más bajos de desgaste. Además, han adaptado sus competencias de tal manera que tienen un buen rendimiento en las tres medidas. Fomentar un enfoque centrado en el usuario podría ser el mejor camino para mejorar el rendimiento empresarial.

Centrarse en los usuarios conduce a un mejor rendimiento empresarial

Datos clave

Las empresas pueden quedarse atascadas con la tecnología más reciente y las tendencias de gestión a medida que buscan mejorar la productividad de los desarrolladores y el rendimiento empresarial. Nuestra investigación demuestra que usar un enfoque centrado en el usuario para desarrollar aplicaciones y servicios es uno de los factores conductores más potentes para el rendimiento general de la empresa. Si quieras mejorar el rendimiento, conoce muy bien a tus usuarios y ajusta e incorpora sus comentarios de forma iterativa.



Introducción

El movimiento DevOps se inició como una forma de fomentar una mejor colaboración entre los equipos de desarrollo y los equipos operativos para ofrecer un mejor valor al usuario. Esta coordinación cosechó éxitos de manera temprana y ha permitido que las ideas y competencias de DevOps se extiendan más allá de estos dos departamentos. Hoy en día, las empresas tecnológicas con un alto rendimiento reconocen la importancia de que todos los equipos estén coordinados para alcanzar los objetivos empresariales.

Hemos estudiado tres características cruciales del enfoque centrado en el usuario:

- Lo bien que entienden los equipos las necesidades de sus usuarios.
- Lo bien que se coordina el equipo para satisfacer las necesidades de los usuarios.
- El uso que se hace de los comentarios de los usuarios al priorizar el trabajo.

Resultados

Hemos descubierto que tener una estrategia de desarrollo de software centrada en los usuarios implica un aumento significativo del rendimiento. Todo son ventajas cuando las empresas anteponen al usuario. Los comentarios de los usuarios ayudan a los equipos a priorizar proyectos y crear productos y servicios que satisfacen sus necesidades. Esta estrategia contribuye a desarrollar una mejor experiencia de usuario y una mayor satisfacción del usuario, además de aumentar los ingresos.

Efectos del enfoque centrado en el usuario

Rendimiento empresarial		Incremento sustancial
Rendimiento del equipo		Incremento sustancial
Rendimiento del envío de software		Incremento pequeño
Rendimiento operativo		Incremento sustancial
Desarrollo basado en troncales		Incremento sustancial
Prácticas de fiabilidad		Incremento sustancial
Integración continua		Incremento sustancial
Entrega continua		Incremento sustancial
Arquitectura con bajo acoplamiento		Incremento sustancial
Desgaste		Reducción pequeña*
Satisfacción laboral		Incremento sustancial
Productividad		Incremento sustancial

* Reducir el desgaste es algo positivo.

¿Qué indican estos resultados?

Centrarse en las necesidades del usuario es un factor destacado para predecir el rendimiento general de la empresa. Para lograr un sólido enfoque en el usuario, se necesitan incentivos, coordinación y formas de trabajar adecuados. Un enfoque en el usuario puede proporcionar información e impulsar mejoras en todas las competencias técnicas, en procesos y culturales que estudiamos en nuestra investigación.

A continuación te explicamos cómo estos resultados pueden afectar a los distintos equipos de tu empresa:

Equipos de desarrollo y entrega de productos

Centrarse en el usuario sirve para confirmar que los equipos de desarrollo y entrega están creando productos adecuados para sus usuarios (y esperamos que lo hagan de un modo sostenible). Esto es lo que justamente hacen los equipos equilibrados. Demuestran lo fuertes que son en el rendimiento del envío, operativo y empresarial gracias a un enfoque claro en las necesidades de los usuarios. Los miembros de estos equipos utilizan a su favor el conocimiento detallado de las necesidades de los usuarios y la capacidad para ajustar los planes en función de los comentarios de los usuarios.

Los resultados muestran que los equipos basados en funciones no han conseguido alcanzar el máximo rendimiento empresarial. Parece que estos equipos priorizan en exceso el rendimiento de la entrega en detrimento del rendimiento empresarial y del bienestar de los empleados en la empresa.

Equipos operativos

Puede que los equipos que se centran en el rendimiento operativo trabajen mucho para optimizar las métricas del sistema, como el uso de la CPU. Pero si no entienden qué espera un usuario

de un servicio, es probable que sigan recibiendo con frecuencia informes de los usuarios que expongan un rendimiento lento. Las prácticas de Site Reliability Engineering (SRE), como identificar los indicadores del nivel de servicio que interesa a los usuarios y establecer objetivos de nivel de servicio que buscan satisfacer al usuario estándar, pueden ayudar a los equipos operativos a adoptar una mentalidad más centrada en el usuario.

» **Equipos de ingeniería de plataformas**

Es posible que los equipos de ingeniería de plataformas adopten un enfoque soñador y piensen "si creamos la plataforma, los clientes vendrán". Tratar a los desarrolladores como usuarios de su plataforma podría ser una estrategia con mejores resultados. Para hacer este cambio de enfoque, es necesario que los equipos de ingeniería de plataformas entiendan cómo trabajan hoy en día los desarrolladores para poder identificar y eliminar los posibles conflictos. Los equipos pueden usar las métricas del rendimiento del envío de software y del rendimiento operativo como señales para monitorizar si los trabajos en la plataforma están sirviendo para que los equipos obtengan mejores resultados.

» **Responsables**

Al crear estructuras de incentivos que premian a los equipos por aportar valor a los usuarios, los responsables ayudan a desarrollar un entorno centrado en el usuario. Sin estas estructuras, es posible que los equipos sientan que únicamente se está midiendo la cantidad de funciones entregadas o la disminución de las interrupciones del servicio. DORA ha estado investigando el papel del liderazgo transformacional¹ y ofrece asesoramiento a los responsables que están listos para mejorar esta función. Para obtener más información, consulta "DevOps Capabilities: Transformational Leadership" en <https://dora.dev/devops-capabilities/cultural/transformational-leadership>

Centrarse en el usuario conduce a un
20%
más de satisfacción laboral

Recursos para empezar

Desarrollar más competencias centradas en el usuario puede ser un factor clave del que dependa el éxito. Nuestras conclusiones del 2023 refuerzan las del 2018,² cuando observamos que las competencias optimizadas de gestión de productos influyen positivamente en el rendimiento del envío de software y el rendimiento empresarial.

Si quieras desarrollar el rendimiento de tu equipo, adopta competencias centradas en el usuario, como pueden ser opiniones de los clientes,³ visibilidad del trabajo en el flujo de valor,⁴ trabajo con lotes pequeños⁵ y experimentación en equipo.⁶

¹Informe State of DevOps 2017. <https://dora.dev/publications/pdf/state-of-devops-2017.pdf>, 12-19

²Informe Accelerate: State of DevOps 2018: Strategies for a New Economy. <https://dora.dev/publications/pdf/state-of-devops-2018.pdf>, 49-51

³"Customer feedback". <https://dora.dev/devops-capabilities/process/customer-feedback/>

⁴"Visibility of work in the value stream". <https://dora.dev/devops-capabilities/process/work-visibility-in-value-stream/>

⁵"Working in small batches". <https://dora.dev/devops-capabilities/process/working-in-small-batches/>

⁶"Team experimentation". <https://dora.dev/devops-capabilities/process/team-experimentation/>

Las competencias técnicas conducen a un mayor rendimiento

Datos clave

Invertir recursos y esfuerzo en la integración continua, una arquitectura con bajo acoplamiento y una velocidad de revisión de código más alta probablemente contribuirá a obtener numerosos resultados beneficiosos, como una mejora del rendimiento empresarial, del rendimiento del equipo, del rendimiento del envío de software y del rendimiento operativo. Esto sucede sin perjuicio (y a menudo con algún beneficio) para el bienestar de los individuos que trabajan en la aplicación o el servicio.

Introducción

En el Resumen ejecutivo hemos explicado las competencias técnicas que hemos estudiado y cómo afectan a diferentes métricas de rendimiento y bienestar. El estudio y la cuantificación del grado en que diversos procesos y competencias técnicas conducen a un mayor rendimiento ha sido siempre un componente central de DORA.

Este año hemos investigado cómo influyen en el rendimiento las siguientes competencias técnicas:

- Inteligencia artificial
- Desarrollo basado en troncales
- Arquitectura con bajo acoplamiento
- Integración continua
- Revisión rápida de código

Nos hemos fijado en cómo influyen en las siguientes métricas de rendimiento:

- Rendimiento del equipo
- Rendimiento empresarial
- Rendimiento del envío de software
- Rendimiento operativo

Además, hemos puesto a prueba el vínculo entre estas competencias y una serie de indicadores para determinar su repercusión sobre las personas que hacen el trabajo:

- Desgaste
- Productividad
- Satisfacción con el trabajo

Resultados

Las competencias técnicas y los procesos que hemos estudiado tienen efectos variados en las métricas de rendimiento clave, aunque su impacto general es positivo.

Competencias técnicas y procesos	Efecto en el rendimiento del equipo	Efecto en el rendimiento empresarial	Efecto en el rendimiento del envío de software	Efecto en el rendimiento operativo				
IA		Ningún efecto demostrado		Incremento pequeño		Reducción pequeña		Reducción sustancial
Integración continua		Incremento pequeño		Incremento pequeño		Incremento pequeño		Ningún efecto demostrado
Velocidad de revisión de código		Incremento pequeño		Ningún efecto demostrado		Incremento sustancial		Incremento sustancial
Arquitectura con bajo acoplamiento		Incremento sustancial		Incremento sustancial		Incremento pequeño		Incremento sustancial
Desarrollo basado en troncales		Incremento pequeño		Incremento pequeño		Incremento pequeño		Reducción pequeña

Los equipos con bajo acoplamiento (o los equipos que tienen una arquitectura con bajo acoplamiento) pueden realizar cambios significativos en sus sistemas sin implicar a otros equipos. Esto permite que los equipos se muevan con mayor rapidez. Cuando los expertos en la materia son cercanos al equipo, pueden revisar el código más rápido porque entienden mejor la repercusión de los cambios. Un diseño con bajo acoplamiento permite al equipo probar, desarrollar y desplegar sin que otros equipos se conviertan en posibles cuellos de botella.

Hasta con el mínimo impacto de nuestros cambios en una arquitectura con bajo acoplamiento, debemos asegurarnos de no estar generando conflictos con el resto de desarrolladores del equipo. Al trabajar con lotes pequeños, los equipos disminuyen las posibilidades de conflicto y se aseguran de que, con cada confirmación, se desarrolla el software y se activan pruebas automatizadas que brindan retroalimentación rápida a los desarrolladores.

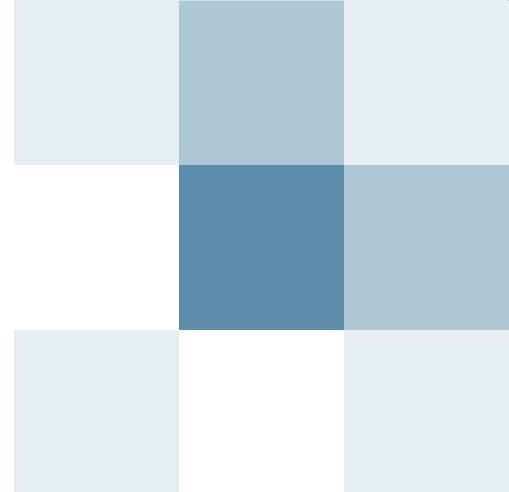
Los equipos con unos tiempos de revisión de código más cortos tienen un rendimiento del envío de software un 50 % mejor. Los procesos de revisión de código eficientes llevan mejoras de código, transferencia de conocimientos, propiedad compartida del código, propiedad del equipo y transparencia.

¿Las revisiones de código son tu cuello de botella? Evalúa tu proceso de revisión de código y su efecto en tu plazo de entrega de los cambios para descubrir oportunidades de mejora. Responde estas preguntas:

- ¿Integras en tu proceso revisiones de código por pares?
- ¿Cuánto tiempo pasa entre la finalización de código y la revisión?
- ¿Cuál es el tamaño medio del lote de tus revisiones de código?
- ¿Cuántos equipos trabajan en tus revisiones?
- ¿Cuántas ubicaciones geográficas distintas participan en tus revisiones?
- ¿Tu equipo mejora la automatización de la calidad del código a partir de las sugerencias de la revisión de código?

Se ha demostrado que, cuanto mayor es el tiempo que se tarda en completar y revisar un código, más repercute de manera negativa en la efectividad del desarrollador y en la calidad del software enviado. La participación de numerosos equipos de diferentes ubicaciones geográficas aumenta la duración del proceso, reduce la implicación e incrementa los costes.¹

En el 2022 descubrimos que las competencias técnicas dependen unas de otras. Mejorar la velocidad de la revisión de código puede contribuir a mejorar otras competencias técnicas, como el mantenimiento del código, la cultura de aprendizaje (transferencia de conocimientos) y el desarrollo de una cultura generativa.²



Una de las ventajas de los equipos con bajo acoplamiento es la mayor rapidez de las revisiones de código, que provoca mejoras significativas en el rendimiento del envío de software y en el rendimiento operativo. Hay varias formas de mejorar la eficiencia de tus revisiones de código. Cuando el código que se revisa solo afecta al ámbito de la arquitectura del equipo, el revisor podrá entender mejor cómo afectará ese código al sistema. Cuanto menor sea la revisión de código, más fácil le resultará al revisor entender las consecuencias del cambio. Trabajar con lotes pequeños mejora el ciclo de comentarios, la eficiencia y el enfoque para el equipo.³ La programación en pareja puede acortar el tiempo de revisión de código independientemente de tu arquitectura actual y tus prácticas de integración.⁴

Además, estas competencias y procesos no demuestran tener un efecto perjudicial en el bienestar de las personas que realizan el trabajo. De hecho, la mayor parte de estos conducen a mejoras en el bienestar de los trabajadores.

¹ Investigating the effectiveness of peer code review in distributed software development based on objective and subjective data. <https://jserd.springeropen.com/articles/10.1186/s40411-018-0058-0>

² Expectations, Outcomes, and Challenges of Modern Code Review. <https://dl.acm.org/doi/10.5555/2486788.2486882>

³ "Working in small batches". <https://dora.dev/devops-capabilities/process/working-in-small-batches/>

⁴ "On Pair Programming". <https://martinfowler.com/articles/on-pair-programming.html>

Competencias técnicas y procesos	Efecto en el desgaste*	Efecto en la satisfacción laboral	Efecto en la productividad			
IA		Reducción pequeña		Incremento pequeño		Incremento pequeño
Integración continua		Ningún efecto		Incremento pequeño		Ningún efecto
Velocidad de revisión de código		Reducción sustancial		Incremento pequeño		Incremento pequeño
Arquitectura con bajo acoplamiento		Reducción sustancial		Incremento sustancial		Incremento sustancial
Desarrollo basado en troncales		Incremento sustancial		Ningún efecto		Ningún efecto

* Fíjate en cómo se invierte el esquema de colores en el caso del desgaste porque reducirlo es algo positivo.

Observamos que el uso de una arquitectura con bajo acoplamiento, integraciones continuas y revisiones de código eficientes permite a los equipos mejorar sus resultados empresariales, conservar y, a veces, mejorar su bienestar.

Cuando los equipos tienen autonomía para perfeccionar y mantener un sistema fiable que aporta valor a los usuarios, mejoran su satisfacción laboral, el rendimiento del equipo y el rendimiento del envío de software. La arquitectura desempeña una función clave en la capacidad que tiene un equipo de enfocarse en el usuario y mejorar su envío de software. Al empezar poco a poco y centrarse en el usuario, los equipos experimentaron mejoras considerables en el desarrollo basado en troncales, la integración continua, la entrega continua y la SRE. Para optimizar tus competencias técnicas, ofrece a los equipos oportunidades para experimentar y fomenta la mejora continua.⁵



⁵ "Team experimentation". <https://dora.dev/devops-capabilities/process/team-experimentation>

Ventajas de la entrega continua

Autor: Dave Farley

El principio fundamental de la entrega continua (CD) es trabajar para que nuestro software siempre se mantenga en un estado implementable. Para conseguirlo, nuestro trabajo debe ser de alta calidad. De esa manera, cuando detectemos un problema, será fácil corregirlo y conseguiremos recuperar su implementación de forma sencilla y rápida.

Para mantener el software en ese estado implementable deseado, debemos trabajar para que los comentarios sean rápidos y podamos recuperarnos muy rápidamente de los fallos.

Como lector del informe de este año, imagino que estas ideas te resultarán conocidas. Las métricas de **Estabilidad** (índice de fallos por cambios y tiempo de recuperación tras un despliegue fallido) se refieren a la calidad, y las métricas de **Rendimiento** (plazo de entrega de los cambios y frecuencia de despliegue) tienen relación con los comentarios y la facilidad para detectar un problema.

En caso de realizar la entrega continua, tus puntuaciones en **Estabilidad y Rendimiento** serán altas. Y si tus puntuaciones en **Estabilidad y Rendimiento** son altas, sería raro que no estuvieras realizando la entrega continua.

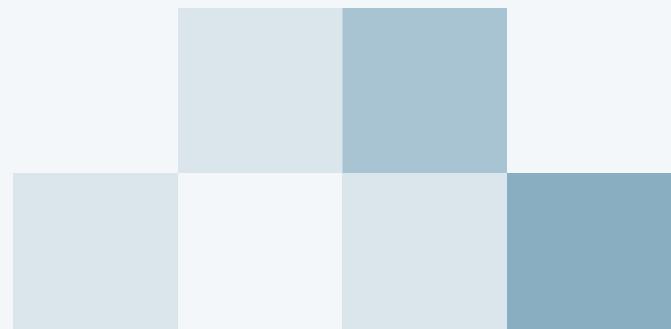
El análisis de este año indaga en cómo las competencias impulsan el rendimiento al buscar mediadores para cada competencia. La entrega continua es la capacidad para implementar cambios de todo tipo bajo demanda de forma rápida, segura y sostenible, así como un importante mediador de muchas competencias técnicas. Dicho de otra

manera, estas competencias funcionan porque crean un entorno que posibilita la entrega continua. Por su parte, la entrega continua proporciona el mecanismo a través del cual estas competencias pueden conducir a un rendimiento del envío de software más sólido.

La implementación es un estándar importante que se debe cumplir en general para el desarrollo de software, y por eso la CD lo resalta. Importa, además, porque es una declaración de calidad subjetiva, aunque definida y aplicable al contexto. El nivel de rigor que determina la implementación puede variar si estamos trabajando en sistemas de seguridad crítica o en un proyecto de software para una pastelería. Pero, en ambos casos, la implementación determina que hemos hecho todo lo que hemos considerado necesario para afirmar que el código está listo y es lo suficientemente bueno y seguro para ponerlo en manos de los usuarios.

Por ello, realizar optimizaciones para que nuestros cambios se puedan implementar también significa optimizar para obtener una definición de la calidad mínima aceptable en un contexto específico de nuestro sistema.

El rendimiento del envío de software es mejor en el caso de los equipos que dan prioridad a obtener comentarios rápidos y de alta calidad y a actuar basándose en ellos.



Me sorprende un poco el hecho de que la integración continua (CI) y el desarrollo basado en troncales no repercutan más en el rendimiento del envío de software. A mí la CI me parece fundamental, por lo que este hecho choca hasta cierto punto con mi visión de la realidad. Sin embargo, es complicado analizar cada elemento por separado. Por ejemplo, ¿cómo podemos alcanzar puntuaciones altas en Rendimiento si nuestro código no se integra? Y ¿cómo podemos estar seguros de que la Estabilidad es alta si no lo hemos comprobado? Para mí, la CI es nuestra manera de saber estas cosas y, por eso, es un mediador clave para el rendimiento del envío de software. ¿Se trata de un problema de interpretación o es algo más importante que va más allá? Resulta, cuanto menos, interesante.

* La mediación es una prueba que evalúa los posibles mecanismos o itinerarios que hay detrás de un efecto. Por ejemplo, puedes decir "los datos confirman la hipótesis de que el efecto del desarrollo basado en troncales sobre el rendimiento del envío de software se produce a través del despliegue continuo (el mediador)". La mediación completa se produce cuando parece que todo el efecto se explica a través del mediador, mientras que la mediación parcial se da cuando solo una parte del efecto se explica a través del mediador.

Competencias técnicas y procesos	Efecto en el rendimiento del envío de software	¿Con mediación gracias a la entrega continua?*
IA	 Reducción pequeña	<input checked="" type="checkbox"/> No
Integración continua	 Incremento pequeño	<input checked="" type="checkbox"/> Sí, totalmente
Velocidad de revisión de código	 Incremento sustancial	<input checked="" type="checkbox"/> Sí, parcialmente
Arquitectura con bajo acoplamiento	 Incremento pequeño	<input checked="" type="checkbox"/> Sí, parcialmente
Desarrollo basado en troncales	 Incremento pequeño	<input checked="" type="checkbox"/> Sí, totalmente

Optimización de procesos empresariales y competencias

Sabemos que la cultura impulsa el éxito. Pero ¿qué impulsa la cultura? Esta es una pregunta interesante que se responde con una respuesta muy típica: depende.

Desde la perspectiva de un profesional, mejorar la forma de trabajar día a día tiene un efecto positivo en elementos de la cultura, como compartir el riesgo, aumentar la cooperación y establecer una seguridad psicológica. Por ejemplo, integrar cambios regularmente en la rama principal del sistema de control de versiones aumenta el intercambio de conocimientos y la colaboración. Si los equipos de seguridad trabajan mano a mano con los desarrolladores y colaboran en las políticas como código, generan más confianza entre los equipos y seguridad en los cambios que se van desplegando.

Desde la perspectiva de liderazgo, la cultura empieza con un reconocimiento y formación sobre la importancia de la cultura. El liderazgo transformacional⁶ puede contribuir a fomentar un entorno irreprochable que anime a experimentar y aprender, y dé confianza y voz a los profesionales. Los ingenieros no solo están para responder a solicitudes de tareas, sino también para solucionar problemas complejos. Para ello, necesitan tener visibilidad del negocio y autonomía para pasar a la acción. En última instancia, la cultura deriva del liderazgo.

Idealmente, los mejores resultados se obtienen al analizar la cultura con una perspectiva tanto descendente como ascendente.

⁶ "Transformational leadership". <https://dora.dev/devops-capabilities/cultural/transformational-leadership>

Inteligencia artificial (IA)

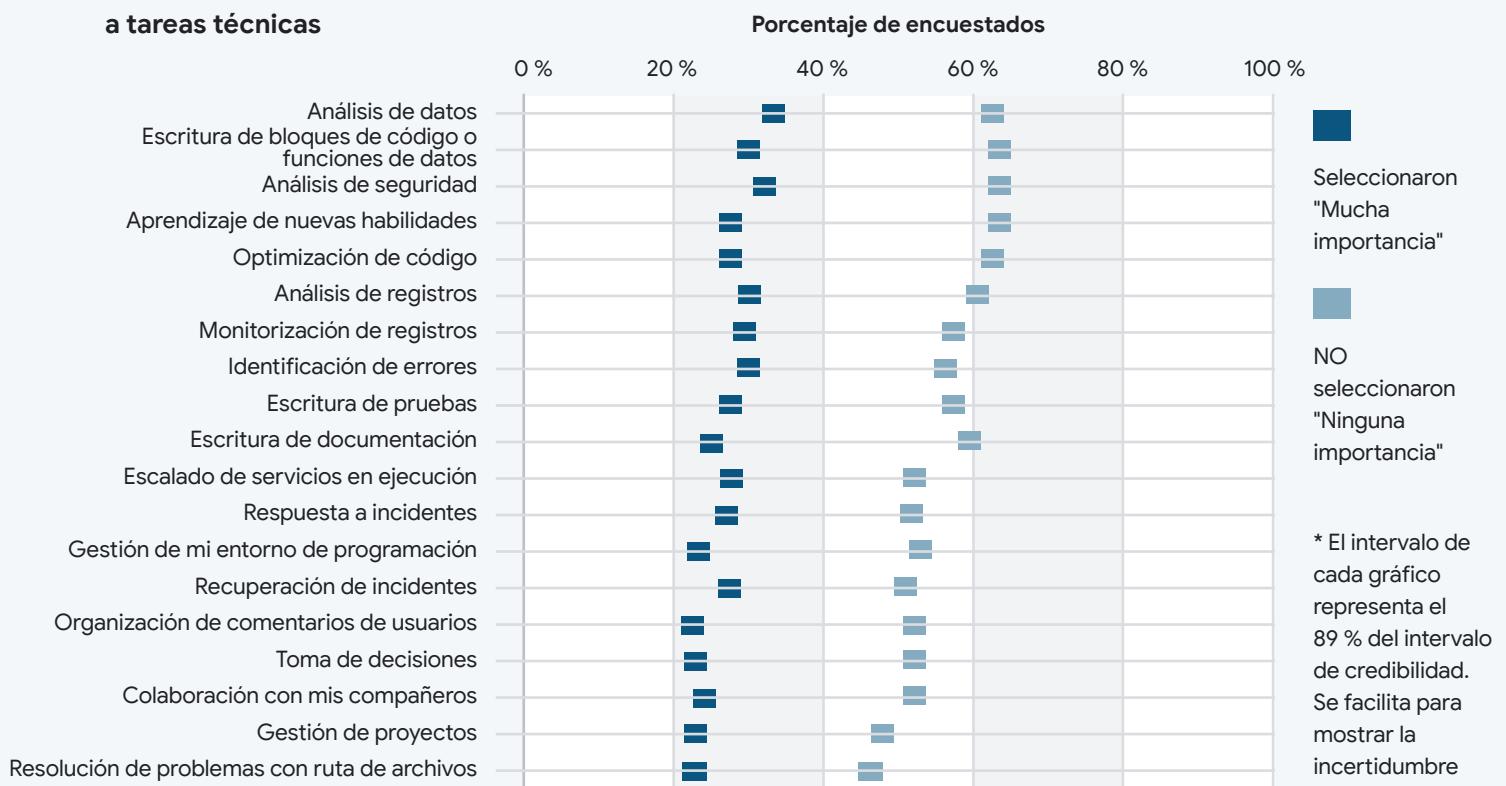
Algunos analistas y técnicos plantean la hipótesis de que la IA hará que los equipos de software rindan más sin que afecte a su bienestar profesional. Hasta la fecha, la evidencia de nuestra encuesta no lo corrobora. Nuestras pruebas sugieren que la IA mejora ligeramente las medidas de bienestar individual (como el desgaste y la satisfacción laboral), pero tiene un efecto neutro o incluso negativo en los resultados a nivel de grupo (como el rendimiento del equipo y el rendimiento del envío de software).

Tenemos la impresión de que la fase inicial en la que se encuentra la adopción de herramientas basadas en IA en las empresas podría explicar esta evidencia desigual. Es posible que algunas empresas grandes estén probando distintas herramientas basadas en IA antes de decidir si utilizarlas o no de manera más generalizada. Existe un entusiasmo general acerca de todo el potencial que tienen las herramientas de desarrollo con IA, como deja claro la mayoría de las personas que incorporan al menos algún tipo de IA en las tareas sobre las que preguntamos. Así se muestra en el gráfico a continuación. Pero anticipamos que las herramientas basadas en IA tardarán un tiempo en alcanzar un uso extendido y coordinado en el sector.

Importancia de la IA

Contribución de la IA a tareas técnicas

Respecto a la aplicación o servicio principal en el que trabajas, ¿qué importancia tiene la inteligencia artificial (IA) en cada una de las siguientes tareas?



Nos interesa mucho ver cómo va creciendo la adopción y el efecto que tiene este crecimiento en las medidas y en los resultados de rendimiento que son importantes para las empresas.

La documentación es fundamental



Datos clave

Es fundamental contar con documentación de calidad. Impulsa una correcta implementación de las competencias técnicas y amplifica su efecto en el rendimiento empresarial. La documentación también tiene un efecto positivo en los resultados, como el rendimiento del equipo, la productividad y la satisfacción laboral. Sin embargo, aunque la calidad de la documentación aumente, eso no implica una mejora en el bienestar de todo el mundo: a medida que esa calidad aumenta, algunos encuestados afirman que los niveles de desgaste son más altos.

Introducción

Este año prestamos mayor atención a la documentación interna, que es la información escrita que las personas de la empresa utilizan en su día a día. Investigamos cómo afecta la documentación a las competencias técnicas y a los resultados clave.

Para medir la calidad de la documentación, valoramos hasta qué punto está actualizada y es fiable, relevante y fácil de encontrar. Después, calculamos la puntuación de toda la experiencia de documentación. No evaluamos la documentación página a página, sino de forma general.

Resultados

La documentación es fundamental: impulsa y amplifica las competencias técnicas

Como ya vimos en el 2021¹ y el 2022², la calidad de la documentación continúa impulsando el éxito de la implementación de cada competencia técnica que estudiamos.

Tal y como muestra la siguiente tabla, la calidad de la documentación también amplifica el efecto de cada competencia técnica en el rendimiento empresarial, que es similar a lo que observamos en el 2022³.

Competencia técnica	Amplificación del efecto en el rendimiento empresarial*
Integración continua	2,4 veces*
Entrega continua	2,7 veces*
Desarrollo basado en troncales	12,8 veces*
Arquitectura con bajo acoplamiento	1,2 veces*
Prácticas de fiabilidad	1,4 veces*
Contribución de la inteligencia artificial	1,5 veces [^]

* Se calcula con la siguiente fórmula:

Efecto de la competencia técnica con documentación de alta calidad

Efecto de la competencia técnica con documentación de baja calidad

¹ Accelerate State of DevOps 2021. <https://dora.dev/publications/pdf/state-of-devops-2021.pdf>

² Informe Accelerate State of DevOps 2022. <https://dora.dev/research/2022/dora-report/2022-dora-accelerate-state-of-devops-report-es.pdf>

³ 2022 State of DevOps Report data deep dive: Documentation is like sunshine. <https://cloud.google.com/blog/products/devops-sre深深理解2022年DevOps报告中的文档>

La documentación influye positivamente en el bienestar y la productividad de las personas y empresas

Además de mejorar las competencias técnicas, descubrimos que una documentación de calidad influye positivamente en el bienestar de las personas: disminuye el desgaste, aumenta la satisfacción con el trabajo e impulsa la productividad. Hemos visto que parte de ese efecto se debe a que la documentación de calidad aumenta el intercambio de conocimientos.

Este resultado no es sorprendente. Es más fácil sacar adelante las cosas cuando sabes cómo hacerlo, y el trabajo frustra menos cuando se intercambian conocimientos.

Aspectos relativos al bienestar	Efecto de la documentación de calidad
Desgaste	 Reducción sustancial
Satisfacción laboral	 Incremento sustancial
Productividad	 Incremento sustancial

La documentación de calidad también hace impulsar resultados clave, que afectan al rendimiento del equipo, al rendimiento empresarial y al rendimiento operativo.

Resultados clave	Efecto de la documentación de calidad
Rendimiento del equipo	 Incremento sustancial
Rendimiento empresarial	 Incremento sustancial
Rendimiento del envío de software	 Ningún efecto*
Rendimiento operativo	 Incremento sustancial

* Nos sigue sorprendiendo no encontrar ningún efecto de la documentación de calidad en el rendimiento del envío de software. Este es el segundo año que lo vemos.

¿Qué se esconde detrás del efecto positivo en los tres resultados clave? Como lector, usar documentación clara es beneficioso. Es posible que el proceso de escritura también sea un factor que debamos tener en cuenta. Para crear documentación de alta calidad, los equipos primero tienen que tomar decisiones sobre los procesos. La documentación puede obligar a los equipos de la empresa a debatir de manera explícita y coordinarse sobre lo que hay que hacer y cómo hacerlo.

La documentación de calidad también funciona como un repositorio de los conocimientos del equipo, aunque lleguen nuevos miembros o algunos se marchen. Ayuda a adaptar el conocimiento a los distintos departamentos de la empresa y a lo largo del tiempo.

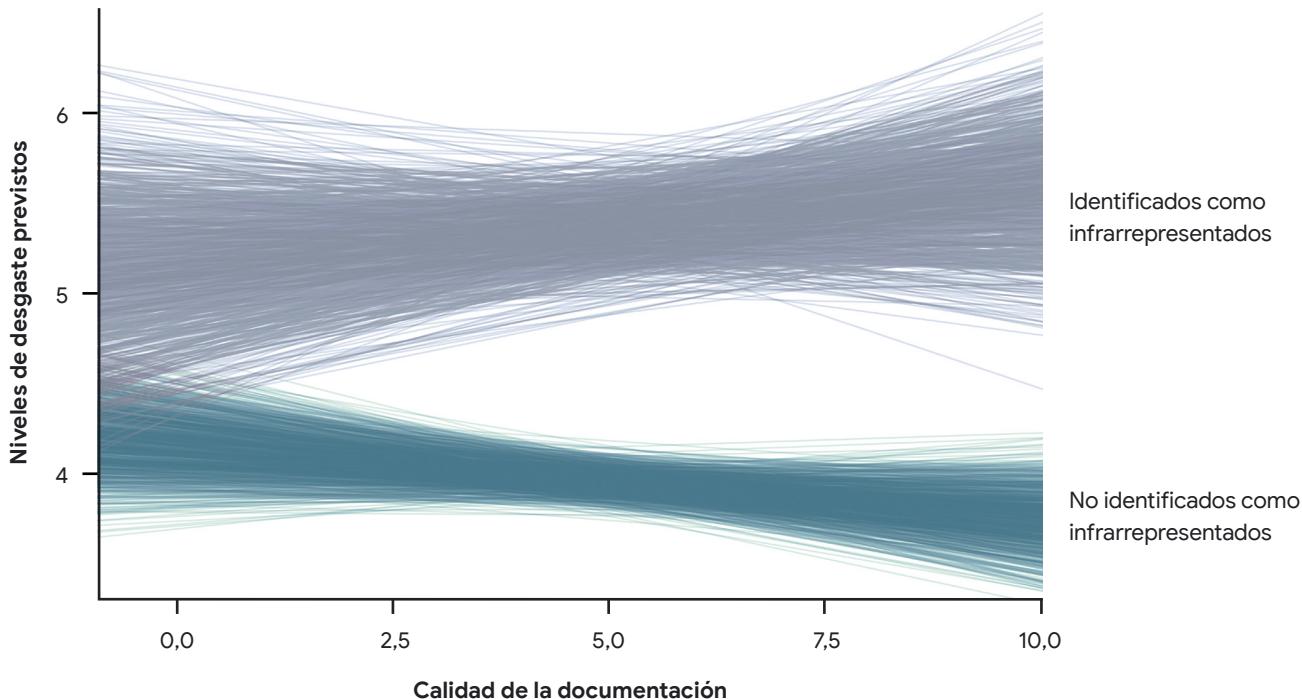
¿La documentación va asociada a una disminución del bienestar de algunas personas?

Hemos detectado una tendencia inesperada al fijarnos en los encuestados que se identifican como pertenecientes a grupos infrarrepresentados. En ese caso, la calidad de la documentación está vinculada a un aumento del desgaste.

Con relación a esta conclusión, nos hemos fijado también en el género y nos ha llamado la atención no encontrar ningún efecto. Los encuestados que se identificaron como hombres, mujeres, así como los que autodeterminaron su género, coincidieron en que el desgaste disminuía de forma significativa gracias a una documentación de alta calidad. Sin embargo, las personas que se identificaron como miembros de grupos infrarrepresentados, independientemente de su identidad de género, experimentaron una tasa más alta de desgaste cuando había documentación de calidad.

El siguiente gráfico muestra una predicción simulada basada en nuestros datos. En la parte inferior observamos un descenso del desgaste para la mayoría de encuestados a medida que la calidad de la documentación aumenta. No obstante, en la parte superior, el desgaste aumenta de manera significativa para las personas que se identifican como miembros de grupos infrarrepresentados.

Este gráfico muestra 1000 líneas simuladas para cada grupo. Las líneas más densas indican que la pendiente es más probable según nuestros datos.



Este resultado es similar para la calidad de la documentación, la cultura generativa y la estabilidad del equipo: cuando estos atributos aumentan, el desgaste también lo hace para las personas que se identifican como miembros de grupos infrarrepresentados. En el caso de la documentación, ¿por qué sucede?

Lleva trabajo crear y mantener una documentación de alta calidad. Es trabajo técnico, con un impacto considerable en las competencias técnicas, la productividad del equipo y el rendimiento empresarial. También es un trabajo cuya importancia y repercusión quizás no siempre se reconozcan como es debido. ¿Las personas que se identifican como infrarrepresentadas hacen buena parte de este trabajo? Y, de ser así, ¿ese trabajo explica el efecto que tiene en el desgaste?

¿Podría causar problemas depender del uso de la documentación? Al aumentar la calidad de la documentación, ¿el intercambio de conocimientos no crece para algunos encuestados? O, si crece, ¿no es suficiente para contrarrestar otros aspectos que contribuyen al desgaste en este grupo?

Es posible que haya algo más que impulse la calidad de la documentación, pero también cree o mantenga el desgaste de los encuestados que se identifican como miembros de grupos infrarrepresentados. Es necesario investigar más.

Parece que importa quién eres en el equipo. Algunas características del lugar de trabajo, como la documentación de calidad, ofrecen beneficios significativos para el equipo y la empresa en general. Pero es posible que también estén relacionadas con resultados negativos para algunos individuos. Lo analizaremos con más detalle en el [Capítulo 8 – Cómo, cuándo y por qué importa ser quien eres](#).

⁴ Accelerate State of DevOps 2021, p. 22. <https://dora.dev/publications/pdf/state-of-devops-2021.pdf#page=22>

Recursos para empezar

Consulta en el informe del 2021 las prácticas que impulsan una documentación de calidad.⁴ Este año también hemos descubierto que la distribución del trabajo (donde se incluyen los procesos formales para distribuir el trabajo de documentación) aumenta de manera considerable la calidad de la documentación.

Hay un montón de recursos y formaciones sobre escritura técnica. Puedes aprender más sobre el tema con estos recursos:

- Society for Technical Communications (stc.org)
- Cursos de escritura técnica para ingenieros (developers.google.com/tech-writing)
- Write the docs (writethedocs.org)

Capítulo 5

La fiabilidad promueve el rendimiento

Datos clave

Unas prácticas de fiabilidad sólidas conducen a una mejora del rendimiento operativo, el rendimiento del equipo y el rendimiento empresarial. Según los datos, los efectos de mejorar estas prácticas siguen un itinerario no lineal. Es decir, a veces puede parecer que las mejoras del rendimiento se estancan cuando una empresa desarrolla unas competencias más sólidas. Sin embargo, mantener el compromiso con estas prácticas a lo largo del tiempo sigue conduciendo a buenos resultados.

Introducción

La fiabilidad es un término ampliamente utilizado en el entorno de operaciones de TI. La definimos como el grado en el que un servicio cumple sus objetivos establecidos con respecto a medidas como la disponibilidad, el rendimiento y la precisión. Una estrategia habitual para lograr resultados de fiabilidad es la SRE, disciplina que se originó en Google (<https://sre.google>) y ahora se pone en práctica en numerosas empresas. SRE prioriza el aprendizaje empírico, la colaboración multidisciplinar, la automatización y el uso de técnicas de medición, como los objetivos de nivel de servicio.

Muchas empresas utilizan prácticas de fiabilidad, pero no las denominan SRE. Entre las formas alternativas de referirse a ellas encontramos: ingeniería de producción, equipos de plataformas, equipo de infraestructuras o TechOps. Para evaluar la magnitud de estas prácticas de la manera más objetiva posible, usamos un lenguaje neutro y descriptivo en nuestra encuesta.

También recopilamos datos sobre los resultados de ingeniería de fiabilidad: es decir, la medida en que los equipos logran alcanzar sus objetivos de fiabilidad. Tanto las prácticas como los resultados de fiabilidad (a los que denominamos **rendimiento operativo**) se incluyen en nuestro modelo predictivo junto con otras competencias.

Prácticas de fiabilidad

Queríamos que los encuestados reflexionaran sobre la fiabilidad y, para ello, les pedimos que pensaran en tres aspectos esenciales de sus operaciones. En primer lugar, ¿disponen de planes de mitigación de sus dependencias? En segundo lugar, ¿prueban frecuentemente sus planes de recuperación tras fallos con interrupciones simuladas, commutaciones por error prácticas o ejercicios de simulación? Por último, cuando no alcanzan sus objetivos de fiabilidad, ¿realizan algún trabajo de mejora o reorganizan sus prioridades y ajustan su trabajo?

Creemos que estas medidas sintetizan el espíritu de un equipo que sigue los principios establecidos de SRE, como "aceptar el riesgo" y "medir la felicidad del usuario". Este tipo de equipo fija un objetivo razonable que está en consonancia con la felicidad del usuario. Después, hacen pruebas para asegurarse de poder cumplir el objetivo en cuestión, aunque, si tienen algún problema, cambian de planes. Vemos esto como indicador de un equipo que "practica SRE" de manera satisfactoria y sin vincular las evaluaciones de los equipos a implementaciones específicas de SRE.

Resultados

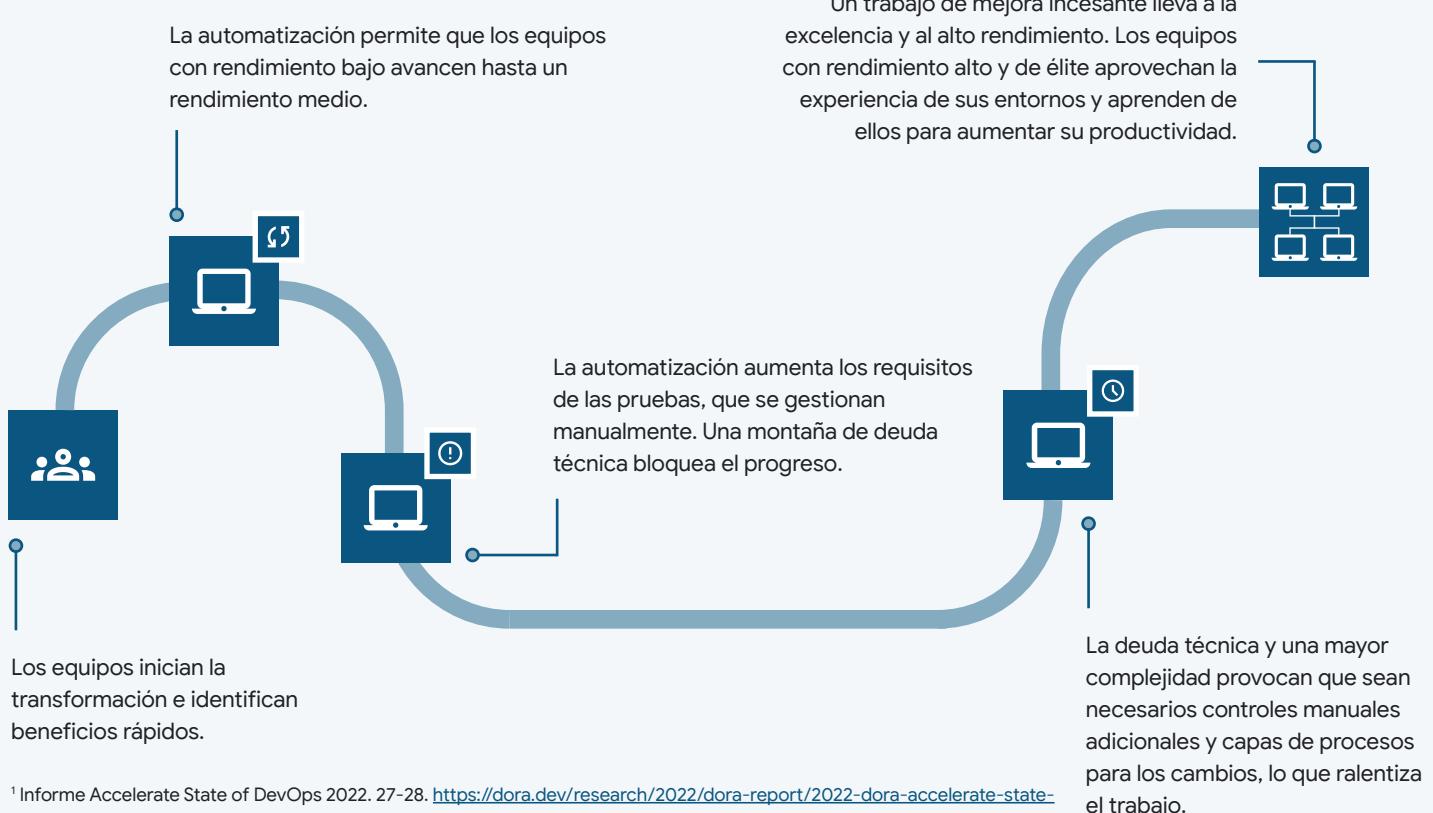
Confirmación de la curva en J de las prácticas de fiabilidad

Desde el 2018, DORA tiene la teoría de que existe una relación no lineal (Figura 1) entre el rendimiento

operativo y prácticas como la automatización. A medida que hemos estudiado más a fondo las prácticas de fiabilidad, hemos observado indicios de este patrón en los datos de la encuesta.

En el 2022 lo medimos directamente. Preguntamos a equipos y observamos que la relación entre las prácticas y los resultados de fiabilidad sigue este tipo de curva no lineal (Figura 2). Esto sugería que los equipos únicamente obtenían mejoras importantes de fiabilidad después de haber adoptado muchas prácticas de fiabilidad.¹ Pero nos dio la impresión de que había algo que se nos escapaba. En la curva del 2022 parecía que SRE solo era para expertos o que no merecía la pena la inversión, lo que chocaba con la experiencia de numerosos equipos SRE. Necesitábamos más datos.

Figura 1: Curva en J hipotética del 2018

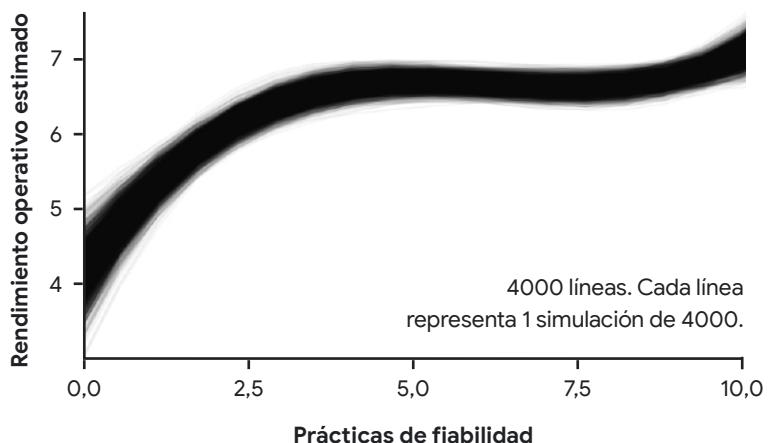


¹ Informe Accelerate State of DevOps 2022. 27-28. <https://dora.dev/research/2022/dora-report/2022-dora-accelerate-state-of-devops-report-es.pdf#page=27>

Figura 2: Curva del 2022



Figura 3: Curva del 2023



En el 2023 hemos podido plantear más preguntas, lo que nos ha ayudado a definir mejor una curva que se ajusta más a nuestras experiencias vividas. La nueva curva se parece más a la curva en J hipotética de transformación descrita en el informe del 2018 (consulta la sección [Metodología](#) para obtener más información sobre cómo realizamos nuestros análisis). Esto nos lleva a pensar que adoptar prácticas de fiabilidad genera unos beneficios iniciales, a lo que sigue un periodo de calma a medida que la complejidad va introduciendo nuevos retos y que, finalmente, se produce otro repunte en el rendimiento operativo. Estos resultados refuerzan lo que hemos visto con numerosos equipos.

La curva es importante por varios motivos:

- Sirve para que las empresas racionalicen y financien la adopción inicial de SRE, aunque no estén buscando unos niveles altísimos de fiabilidad o no tengan intención de invertir de manera significativa en SRE. Incluso pequeñas prácticas de fiabilidad pueden conllevar mejoras de rendimiento operativo, que tienen efectos positivos adicionales en el rendimiento del equipo y en el rendimiento empresarial.

- Prepara a las empresas que están pensando en invertir considerablemente en fiabilidad para que puedan aguantar durante los períodos de calma. Es tentador esperar resultados positivos lineales por la inversión a largo plazo en SRE, pero los datos indican que no sucede así. Cuando los equipos saben con antelación que esta curva no es lineal, pueden decidir si hacer o no la inversión y anticiparse para no abandonar antes de aprovechar todas sus ventajas.
- Es probable que este tipo de cambios requiera una transformación cultural.² Hemos descubierto que el éxito se consigue al combinar cambios ascendentes y descendentes. Los equipos pueden adoptar prácticas de fiabilidad y recibir recompensas inmediatas. Después, esas ventajas se pueden mostrar a los demás equipos con el refuerzo y el incentivo de los responsables. Estos incentivos y programas estructurados se pueden diseñar teniendo en cuenta la curva en J.

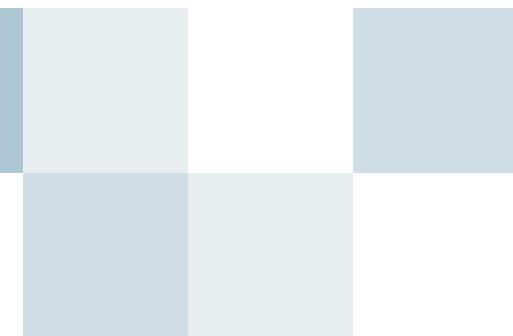
² "How to transform" - <https://dora.dev/devops-capabilities/cultural/devops-culture-transform/>

Prácticas de fiabilidad y bienestar

Las prácticas operativas tradicionales son muy reactivas y suelen tener más relación con el estado del sistema técnico que con la felicidad de sus usuarios. Las alertas de guardia por incidentes que no repercuten en las experiencias de los usuarios, las tareas manuales repetitivas, el miedo a cometer errores y otras experiencias similares llevan al desgaste y al malestar de los miembros del equipo.

En los equipos que aprovechan las prácticas de fiabilidad vemos justo lo contrario. Los equipos registran más productividad y satisfacción con el trabajo y unos niveles más bajos de desgaste que los compañeros que no hacen uso de estas prácticas. Suponemos que estas mejoras en el bienestar se deben a algunas prácticas de SRE publicadas:

- Disminución de labores manuales³
- Análisis post mortem sin culpa⁴
- Autonomía del equipo⁵
- Escalado sublineal de los equipos⁶



³Beyer, Betsy et al. Site Reliability Engineering: How Google Runs Production Systems (O'Reilly, 2016), 49–54. <https://sre.google/sre-book/eliminating-toil/>

⁴Ibid, <https://sre.google/sre-book/postmortem-culture/>

⁵Beyer, Betsy et al. The Site Reliability Workbook (O'Reilly, 2018). <https://sre.google/workbook/team-lifecycles/>

⁶Brookbank, James y McGhee, Steve. Enterprise Roadmap to SRE (O'Reilly, 2022), 11. <https://sre.google/resources/practices-and-processes/enterprise-roadmap-to-sre/>

Rendimiento operativo

También pedimos a los encuestados que describieran el rendimiento operativo de su servicio. En primer lugar, les preguntamos con qué frecuencia escuchan directamente de los usuarios que no están satisfechos con la fiabilidad de su servicio. Luego les preguntamos con qué frecuencia su servicio no está disponible, es lento o funciona incorrectamente.

» Las prácticas de fiabilidad amplifican el rendimiento empresarial y del equipo mediante el rendimiento operativo

Al adoptar prácticas de fiabilidad, los equipos mejoran su rendimiento operativo. Hemos observado que si una empresa es capaz de operar su flota de producción de manera efectiva, se *amplifican* otros resultados. Si los resultados son altos, crecerán aún más con las prácticas de fiabilidad. Si los resultados son bajos, las prácticas de fiabilidad no servirán para nada y simplemente se quedarán como están.

Aunque los sistemas sean fiables, sigue siendo necesario que las competencias en software adecuadas se proporcionen a los clientes de manera efectiva. Esto tiene todo el sentido porque SRE nunca se concibió para funcionar de forma aislada. Cumplir los objetivos de fiabilidad es una métrica clave de éxito para los equipos de SRE, y se refleja en el rendimiento operativo. Es probable que el uso de prácticas de fiabilidad tenga otras ventajas, pero los datos sugieren que la más importante es el efecto en el rendimiento operativo. Además, un mayor rendimiento operativo ofrece otras ventajas más allá del estado del servicio. De hecho, observamos indicios de que el uso de prácticas de fiabilidad repercute en un mayor bienestar para los profesionales.

» El rendimiento operativo afecta al bienestar

En el sector parece que los servicios con alta fiabilidad afectan negativamente al bienestar de los operadores de servicios, por ejemplo, por las guardias

o el mantenimiento de emergencia fuera del horario laboral. Sin embargo, hemos descubierto que un rendimiento operativo elevado realmente produce menos desgaste, mejora la productividad y aumenta la satisfacción con el trabajo. Lo que describimos está en sintonía con el principio de SRE de disminuir la labor manual;⁷ automatizar las partes manuales de las operaciones resulta gratificante para los individuos y conlleva una carga de trabajo constante menor para el equipo.

El rendimiento operativo amplifica el rendimiento del equipo y el de la empresa

Hemos visto que el rendimiento operativo tiene un efecto positivo significativo tanto en el rendimiento del equipo como en el de la empresa. Esto no debería sorprender a los seguidores del movimiento DevOps. Si todo funciona eficazmente, los equipos pueden alcanzar más metas, lo que permite a las empresas prosperar.

El rendimiento operativo amplifica el rendimiento del envío de software

Aunque el rendimiento del envío de software puede mejorar tanto el rendimiento del equipo como el rendimiento empresarial, ambos se ven reforzados enormemente por el rendimiento operativo. Además, los equipos de envío de software con gran rendimiento no lograrán altos niveles de rendimiento de equipo ni empresarial si no alcanzan también un alto rendimiento operativo. Ambos son necesarios. De hecho, los equipos que mejoran el rendimiento del envío de software sin los niveles similares de rendimiento operativo terminan teniendo peores resultados a nivel de empresa. Por tanto, si puedes escribir software increíble rápidamente, pero falla al pasar a producción y no cumple las expectativas de la audiencia, el mercado no lo recompensará.

¿Qué falta? ¿Qué es lo siguiente?

Creemos que hay más métricas que pueden ayudarnos a entender estas interacciones. Por ejemplo, una pregunta habitual este año ha sido cómo encaja la gestión del coste en estas competencias y resultados. Algunas empresas dependen más de los costes que otras, lo que repercute en su manera de planificar y tomar decisiones. De forma similar, tenemos la teoría de que en culturas muy colaborativas pueden surgir prácticas de fiabilidad,⁸ aunque no se estuvieran buscando ni planeando explícitamente. Queremos llegar a entender bien cómo desarrollan los equipos sus prácticas de operaciones de TI y cómo afecta ese desarrollo a la fiabilidad del sistema, al rendimiento del equipo y al bienestar.

Sobre todo, queremos conocer tu opinión. Únete a nosotros y a otros profesionales en [DORA.community](https://dora.community/).⁹ SRE sigue siendo un campo novedoso. Tiene un efecto diferente en cada empresa que adopta prácticas de fiabilidad, o en esas empresas que se dan cuenta de que llevan todo este tiempo haciendo SRE. Los cambios son lentos y queremos hacer mediciones coherentes para ver lo que se ha avanzado con el paso del tiempo. Como comunidad, podemos compartir lo que funciona y así hacernos mejores unos a otros.

⁷ Beyer, Betsy et al. Site Reliability Engineering: How Google Runs Production Systems (O'Reilly, 2016), 49–54. <https://sre.google/sre-book/eliminating-toil/>

⁸ Brookbank, James y McGhee, Steve. Enterprise Roadmap to SRE (O'Reilly, 2022), 5. <https://sre.google/resources/practices-and-processes/enterprise-roadmap-to-sre/>

⁹ DORA Community, <https://dora.community/>

Cómo hace SRE Google

En Google, SRE lleva dos décadas de desarrollo dentro de una empresa en crecimiento con muchos recursos financieros. El auge de SRE se produjo por la necesidad de habilitar el hipercrecimiento de la Búsqueda de Google y Google Ads sin invertir mucho dinero. Desde el principio, la naturaleza inmediata de estos productos puso de manifiesto la necesidad de una fiabilidad alta: un error pasajero en la búsqueda o los anuncios implica perder un cliente al instante, y no hay segundo intento que valga. Un anuncio dinámico debe calcularse en milisegundos. Una búsqueda lenta ensucia la imagen de la marca Google.

Al mismo tiempo, SRE se estaba desarrollando al abrigo de un nuevo tipo de empresa: una compañía guiada por la ingeniería ascendente que prefiere desarrollar a comprar. Google se valió de un equipo de ingenieros de fiabilidad del sitio con un alto nivel en materias de informática, como los sistemas distribuidos y el diseño de compiladores. Ese ADN cultural creó un ambiente propicio para que surgiera y prosperara la SRE. El equipo de SRE no solo recibió las llaves de la producción, sino también el encargo de buscar formas nuevas y vanguardistas de escalar los sistemas.

¿Pero cómo escaló verticalmente SRE con el tiempo? En una palabra: sublinealmente. Es decir, Google no podía duplicar el número de ingenieros de SRE cada vez que duplicaba su tamaño. Al ritmo al que escalaban verticalmente los primeros productos de Google para satisfacer la demanda mundial (y con la introducción de nuevos productos, como Gmail, Google Maps, Android, YouTube y Google Cloud), no era posible escalar el número de estos nuevos y escasos ingenieros de SRE con la misma rapidez a la que crecía la base de clientes.

Empezó a decirse lo siguiente:

- SRE no crecerá de forma lineal con respecto a la cantidad de usuarios.
- SRE no crecerá de forma lineal con respecto a la cantidad de servidores.
- SRE no crecerá de forma lineal con respecto a la cantidad de grupos.
- SRE no crecerá de forma lineal con respecto a la cantidad de servicios.

Se desarrolló una estructura de gestión para permitir que se mantuviese este modelo de crecimiento restringido. Los ingenieros de SRE se coordinaron y cooperaron unos con otros mientras seguían de cerca el trabajo de los equipos de desarrollo de productos con los que colaboraban. Los ingenieros de SRE informaban a sus superiores través de su propia cadena de gestión, con Ben Treynor-Sloss¹⁰ a la cabeza. Estos equipos se segmentaron en áreas de producto que se ajustaban exactamente a las áreas de producto del desarrollo de productos ("desarrolladores"). Equipo de ingenieros de SRE trabajaron junto con sus equipos de desarrollo para decidir la mejor forma de emplear a los ingenieros de SRE con los que contaban. Los equipos de desarrollo podían

¹⁰ Benjamin Treynor Sloss, vicepresidente de Ingeniería. Site Reliability Engineering: How Google Runs Production Systems (O'Reilly, 2016). <https://sre.google/sre-book/part-1-introduction/>, <https://sre.google/sre-book/introduction/#id-2opuzSjFr>

financiar directamente a los nuevos SRE. No todos los equipos de desarrollo crearon productos para los usuarios; muchos eran equipos de infraestructura compartida, como Bigtable¹¹ (almacenamiento de datos estructurados), Borg¹² (planificación de cómputo) y Colossus¹³ (almacenamiento distribuido). Estos servicios de infraestructura compartida permitieron escalar los equipos de productos para clientes sin un equipo específico de SRE.

Al disponer de estos equipos dentro de la propia empresa, los equipos de desarrollo podían mantener un proceso coherente de contratación y promoción. Los equipos de SRE tendían a ser, como mínimo, diez veces más pequeños que sus equipos de desarrollo, por lo que era importante confirmar la autonomía de los equipos de SRE para que no se vieran arrastrados hacia ninguna dirección que contraviniere los principios de esta ingeniería.

Los equipos de SRE desarrollaban sus propios productos internos con gestores de productos internos. Los clientes de los productos de SRE eran otros equipos que buscaban mejorar la ejecución de la producción. Los equipos de SRE desarrollaron productos relacionados con la respuesta a incidentes, su monitorización y observabilidad, la gestión de lanzamientos, la planificación de capacidades y la resolución de conflictos.

El proceso de crecimiento de los equipos de SRE siempre tenía en cuenta la salud y la sostenibilidad del equipo. Por ejemplo, los nuevos equipos de SRE de guardia tenían un tamaño mínimo de 12 miembros: 2 lugares de trabajo con 6 personas cada uno. Esto permitía cubrir varias zonas horarias y, al tener suficientes personas en un equipo, era posible conciliar la vida laboral y personal sin que nadie tuviera que sufrir desgaste.

SRE sigue adaptándose hoy en día. No todos los equipos han seguido con precisión los modelos que se describen arriba, algunos incluso han optado por rutas alternativas. A las organizaciones grandes como SRE en Google también les afectan las condiciones del mercado, como a cualquier otro grupo de gran tamaño, por lo que la flexibilidad es importante. Pero, ante todo, los equipos de SRE se mantienen fieles a sus principios: aceptar el riesgo, medir los niveles de servicio, eliminar la labor manual, aceptar la automatización y buscar la simplicidad.



¹¹ Chang, Fay, et al. Bigtable: A Distributed Storage System for Structured Data, 7th USENIX Symposium on Operating Systems Design and Implementation (OSDI), {USENIX} (2006), págs. 205-218, <https://research.google/pubs/pub27898/>

¹² Verma, Abhishek et al. Large-scale cluster management at Google with Borg. Proceedings of the European Conference on Computer Systems (EuroSys), ACM, Burdeos, Francia (2015), <https://research.google/pubs/pub43438/>

¹³ Hildebrand, Dean et al. Colossus under the hood: a peek into Google's scalable storage system. 19 de abril del 2021 - <https://cloud.google.com/blog/products/storage-data-transfer/a-peek-behind-colossus-googles-file-system>

Capítulo 6

Una infraestructura flexible es crucial para el éxito

Datos clave

La infraestructura favorece el rendimiento del equipo, el rendimiento empresarial, el rendimiento operativo y el rendimiento del envío de software. El cloud computing es esencial para facilitar una infraestructura flexible, aunque el beneficio no se materializa automáticamente, ya que, según nuestros datos, lo importante es cómo se usa la nube.



Introducción

Durante la mayor parte de la investigación de DORA, hemos preguntado a los profesionales acerca de su infraestructura y nos hemos centrado en las características clave de cloud computing, tal y como define el Instituto Nacional de Normas y Tecnología (NIST):¹

- Autoservicio bajo demanda
- Amplio acceso a la red
- Agrupamiento de recursos
- Elasticidad rápida
- Servicio medido

Hemos observado reiteradamente que estas cinco características conducen a una mejora del rendimiento empresarial y del rendimiento del envío de software. Este año queríamos ver si el uso de cloud computing conducía a una infraestructura más flexible.

¹Publicación especial 800-145 del NIST: "The NIST Definition of Cloud Computing".

	Porcentaje
Multinube	19,6 %
Nube pública	51,5 %
Nube híbrida	33,6 %
On-premise	19,8 %
Servidor local	3,2 %
Otros	2,5 %

* Los encuestados pudieron seleccionar varias respuestas.

Resultados

Se ha vuelto a confirmar esta conclusión: cómo usa un equipo la nube es un factor de predicción del rendimiento más importante que el simple hecho de que ese equipo use la nube. Aunque el uso de la nube puede ser un potente facilitador, no produce beneficios de forma automática. De hecho, hay indicadores claros de que la nube pública conduce a un *menor* rendimiento operativo y de software *a menos* que los equipos utilicen una infraestructura flexible. Esta conclusión además promueve la idea de que simplemente "migrar mediante lift-and-shift" (la acción de mover cargas de trabajo de un centro de datos a la nube) no es beneficiosa y podría ser perjudicial.

El uso de cloud computing se asocia con una reducción esencial del desgaste y un incremento esencial de la satisfacción laboral y la productividad.

Entornos de computación

En la tabla anterior se muestra dónde han dicho los encuestados que se está ejecutando su aplicación o servicio principal.



Tipo de nube	Rendimiento empresarial	Rendimiento del equipo	Rendimiento del envío de software	Rendimiento operativo
Privada		No se aprecia ningún efecto		No se aprecia ningún efecto
Pública				Incremento sustancial asociado al uso de cloud computing
Híbrida		Incremento muy sustancial asociado al uso de cloud computing		Reducción sustancial asociada al uso de cloud computing
Multi				Reducción sustancial asociada al uso de cloud computing

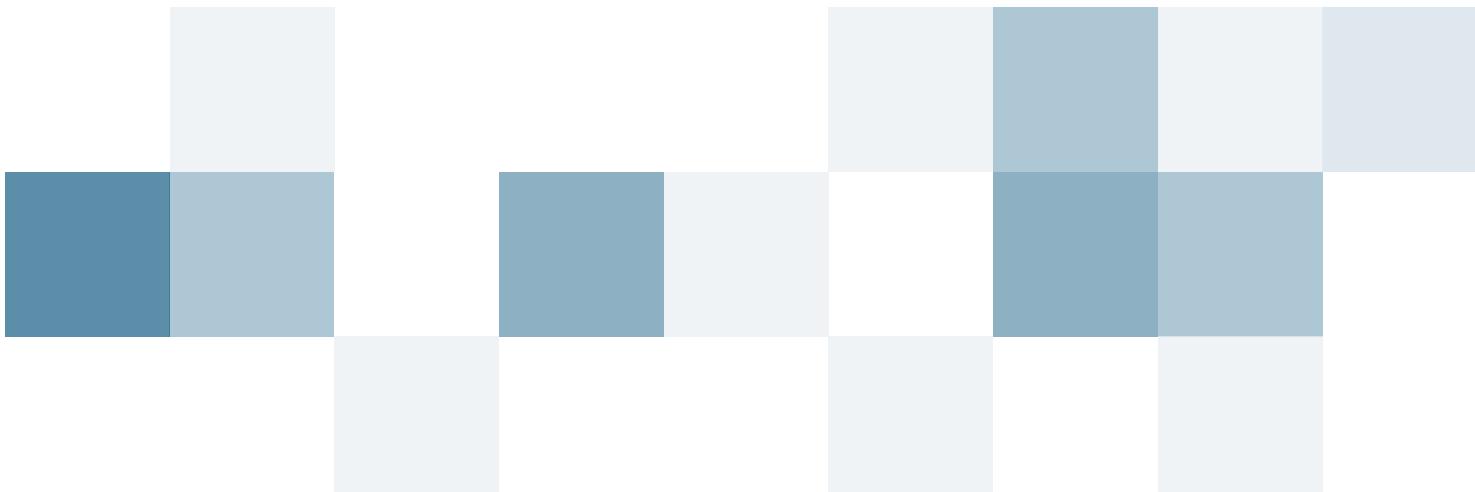
Simplemente "usar la nube" tiene resultados desiguales

Como muestran los resultados de la tabla, simplemente "usar la nube" tiene efectos neutros o negativos en el envío de software y el rendimiento operativo. Este efecto neutro o negativo probablemente se deba a que los profesionales han iniciado su camino en la nube y ahora se enfrentan a trabajar en un entorno nuevo con nuevas herramientas y haciendo *algunas* tareas de otra forma. A menudo, las empresas utilizan la nube de la misma manera que han usado sus centros de datos, pero con las complejidades y la carga cognitiva

añadidas de un entorno nuevo. Si no se adaptan a este nuevo entorno, no mejora el envío de software ni el rendimiento operativo, sino todo lo contrario.

La única excepción a este resultado es el del rendimiento operativo en el contexto de una nube privada.

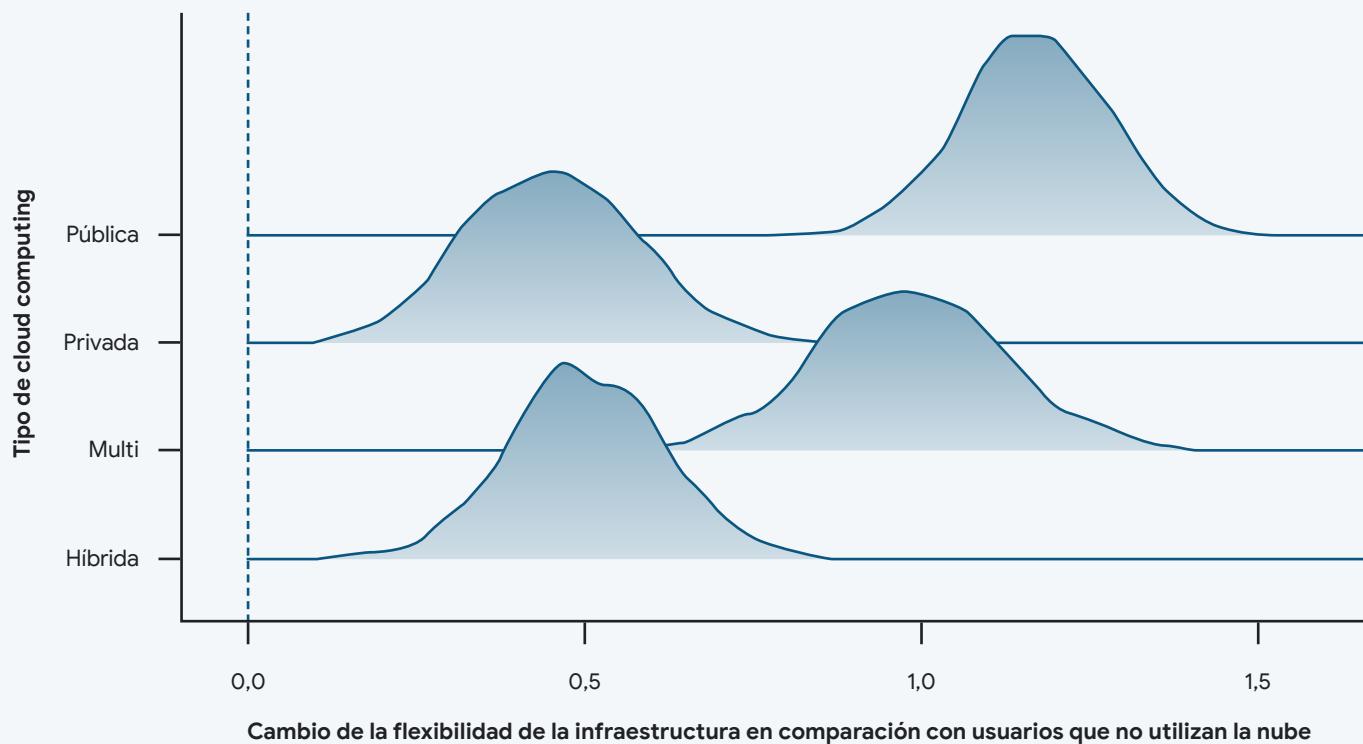
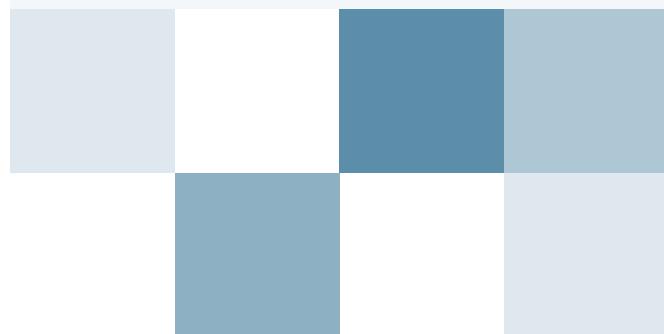
Lo que sí mejora el envío de software y el rendimiento operativo es la **infraestructura flexible**, de la que hablaremos enseguida.



La infraestructura en la nube facilita la flexibilidad

El uso de una nube pública aumenta un 22 % la flexibilidad de la infraestructura en comparación con no usar la nube. Utilizar varios tipos de nube también produce un incremento, aunque inferior al de una nube pública. Cabría preguntarse por qué. Nuestros datos muestran que una infraestructura flexible, a menudo facilitada por cloud computing, tiene mayor repercusión que simplemente usar una plataforma en la nube. Para la mayoría, la nube supone hacer las cosas de una forma nueva, y dominarla lleva su tiempo. Cada plataforma en la nube es diferente y eso significa que, a medida que aumentan las plataformas en la nube, incrementa también la carga cognitiva que se requiere para gestionar cada plataforma correctamente.

Usar una nube pública aumenta la flexibilidad de la infraestructura en comparación con no usar la nube.



Las infraestructuras flexibles conducen a un mayor rendimiento en los resultados clave

Competencia	Rendimiento empresarial	Rendimiento del equipo	Rendimiento del envío de software	Rendimiento operativo
Infraestructura flexible	Incremento sustancial asociado a una infraestructura más flexible	Incremento sustancial asociado a una infraestructura más flexible	Incremento sustancial asociado a una infraestructura más flexible	Incremento asociado a una infraestructura más flexible

Es importante reconocer que la infraestructura flexible hace impulsar el rendimiento empresarial, el rendimiento del equipo, el rendimiento del envío de software y el rendimiento operativo. Muchas empresas deciden migrar la infraestructura a la nube mediante lift and shift, y este puede ser un primer paso excelente, pero no es más que el principio del camino. Si finalmente decides migrar una parte de tus cargas de trabajo mediante lift and shift, el siguiente paso será modernizarlas con refactorización para usar la infraestructura flexible.

El cloud computing tiene un efecto positivo en los resultados clave gracias a la infraestructura flexible

Tipo de nube con infraestructura flexible	Rendimiento empresarial	Rendimiento del equipo	Rendimiento del envío de software	Rendimiento operativo
Privada	Mediado parcialmente por la infraestructura flexible			
Pública	Mediado totalmente por la infraestructura flexible	Mediado totalmente por la infraestructura flexible	Mediado parcialmente por la infraestructura flexible	Mediado parcialmente por la infraestructura flexible
Híbrida	Mediado parcialmente por la infraestructura flexible	Mediado totalmente por la infraestructura flexible	Mediado parcialmente por la infraestructura flexible	Mediado totalmente por la infraestructura flexible
Multi	Mediado parcialmente por la infraestructura flexible			

Las plataformas de cloud computing, cuando se utilizan de manera que sacan el máximo partido a las características de la infraestructura flexible, conllevan un efecto positivo en el envío de software y en el rendimiento operativo. La diferencia que observamos en el efecto que tiene respaldo lo que la mayoría de profesionales y responsables ya sabían: el éxito no se consigue simplemente con cambiar las cargas de trabajo de un centro de datos a la nube. La clave está en aprovechar la infraestructura flexible que habilita la nube.

Tipo de infraestructura	Resultados
Nube acoplada con infraestructura flexible	
Nube con flexibilidad	

Si quieras aumentar al máximo tu potencial de beneficio, debes darle una vuelta a cómo desarrollas, pruebas, despliegas y monitorizas tus aplicaciones. Para ello, deberías sacar partido a las cinco características de cloud computing: autoservicio bajo demanda, amplio acceso a la red, agrupamiento de recursos, elasticidad rápida y servicio medido.



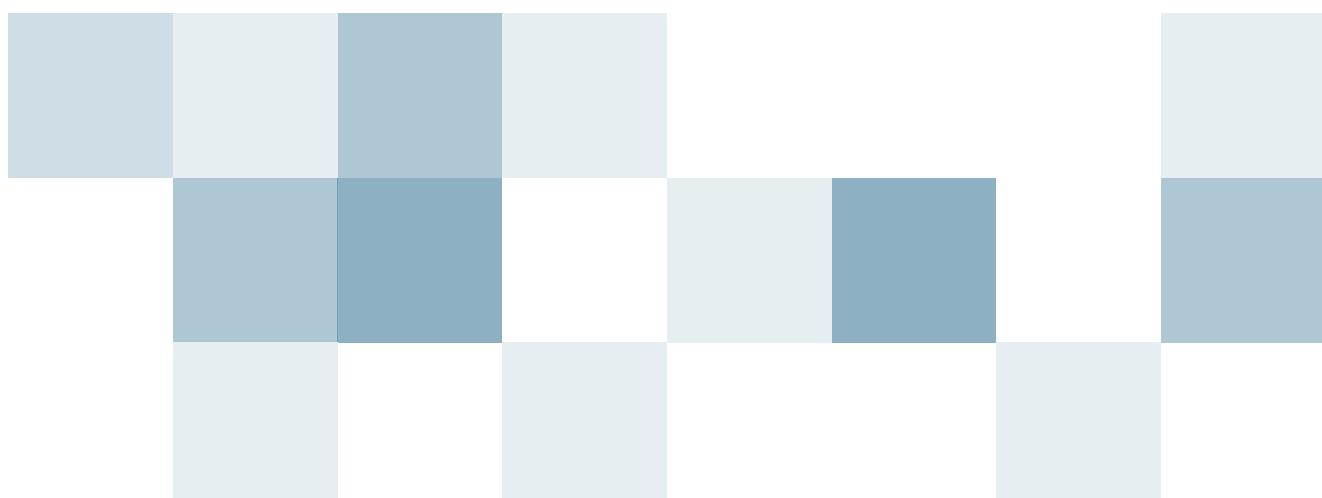
El cloud computing mejora el bienestar

Tipo de nube	Desgaste*	Satisfacción laboral	Productividad	
Privada		No se aprecia ningún efecto		
Pública		Reducción muy sustancial asociada al uso de cloud computing		Incremento sustancial asociado al uso de cloud computing
Híbrida		No se aprecia ningún efecto		
Multi		No se aprecia ningún efecto		

* Fijate en cómo se invierte el esquema de colores en el caso del desgaste porque reducirlo es algo positivo.

Los datos muestran que el cloud computing es enormemente beneficioso para el bienestar del empleado. Observamos que tanto la satisfacción laboral como la productividad aumentan sustancialmente, mientras que el efecto es neutro o positivo en el caso del desgaste. Es decir, el uso de la nube no perjudica el bienestar, aunque el cloud computing suponga una carga cognitiva mayor, el aprendizaje de nuevas herramientas y nuevas formas de trabajar.

Como profesionales, planteamos hipótesis sobre por qué sucede esto. A los ingenieros les gusta aprender y solucionar problemas, y disfrutan trabajando en un entorno con características de computación flexibles. Aprender nuevas tecnologías no solo les divierte, sino que es una forma genial de desarrollar su carrera profesional. Los ingenieros se sienten más satisfechos cuando a sus empresas les va bien.



Capítulo 7

Nada de esto es posible si no se invierte en la cultura

Datos clave

La cultura es un factor clave para el bienestar del empleado y el rendimiento empresarial. Una cultura sana puede ayudar a disminuir el desgaste y aumentar la productividad y la satisfacción laboral. También contribuye a mejorar de forma significativa el rendimiento empresarial, del envío de software, operativo y del equipo. Una cultura empresarial sana puede ayudar a los equipos a implementar con éxito competencias técnicas asociadas a la mejora de resultados.



Introducción

Es difícil definir el concepto de cultura, así que vamos a centrarnos en indicadores relacionados con la experiencia de las personas en el trabajo. Utilizamos la tipología de la cultura empresarial de Westrum¹ porque esta siempre ha pronosticado el rendimiento eficazmente. Este año observamos que las empresas con una cultura que Westrum definiría como generativa siguen teniendo un buen rendimiento.

En la siguiente tabla se enumeran aspectos que creemos que contribuyen a la cultura del equipo y la empresa.

Aspecto	Definición
Cultura empresarial de Westrum	La forma habitual que tienen las empresas de responder a problemas y oportunidades. Existen tres tipos de cultura: generativa, burocrática y patológica.
Estabilidad empresarial	Lo estable o inestable que es un ambiente para los empleados.
Seguridad laboral	La frecuencia con la que los empleados se preocupan por su seguridad en el trabajo.
Flexibilidad	Cómo, dónde y cuándo se trabaja.
Intercambio de conocimientos	La forma de difundir ideas e información en una empresa. Los miembros del equipo responden a las preguntas una vez, y la información pasa a estar disponible para los demás. Las personas no tienen que esperar para recibir respuestas.
Enfoque centrado en el usuario	Enfoque en el usuario final al desarrollar software y una mejor comprensión de las necesidades y objetivos de los usuarios. Los comentarios de los usuarios se utilizan para mejorar los productos y servicios.
Distribución del trabajo	Procesos formales que ayudan a los equipos a distribuir tareas tediosas de manera equitativa entre sus miembros.

Las líneas que separan los aspectos culturales, las competencias en procesos y las competencias técnicas no siempre están claras. Creemos que la cultura surge de las prácticas y las prácticas, de la cultura. Abordaremos este tema más adelante cuando hablemos de las conclusiones.

¹<http://bmj.co/1BRGh5q>

¿Qué hemos descubierto y qué significa?

Una cultura sana mejora los resultados clave.

En general, una cultura sana tiene un efecto positivo en todos los resultados clave. Las conclusiones de años anteriores se repiten: una cultura generativa impulsa el rendimiento empresarial, el rendimiento del envío de software y el rendimiento operativo. También impulsa la nueva métrica de rendimiento de este año, el rendimiento del equipo.

Además, hemos descubierto que usar una estrategia centrada en el usuario para desarrollar software aumenta significativamente el rendimiento, lo cual merece la pena destacar. Todo son ventajas cuando las empresas anteponen al usuario. Los comentarios de los usuarios ayudan a los equipos a priorizar proyectos y crear productos y servicios que satisfacen sus necesidades. Esto contribuye a desarrollar una mejor experiencia de usuario y una mayor satisfacción del usuario, además de aumentar los ingresos.

También hemos medido la distribución del trabajo entre los equipos para evaluar la salud de una cultura empresarial, y hemos descubierto que una distribución equitativa del trabajo beneficia al equipo y al rendimiento de la empresa. Sin embargo, hemos observado que la distribución equitativa del trabajo se asocia a un rendimiento del envío de software menor. Puede que los procesos formales relacionados con la distribución del trabajo ralenticen la ejecución de tareas tediosas que forman parte del flujo de procesamiento del envío de software. También es posible que los procesos formales influyan en quién debe asumir una tarea determinada dentro del equipo.

Los equipos con culturas generativas tienen un rendimiento empresarial un

30%

mayor que los equipos sin ellas

Otro resultado aparentemente incongruente es que observamos que la estabilidad de la empresa sufre una pequeña disminución del rendimiento del envío de software, que resulta llamativa. Una posible explicación sería que las empresas más longevas (y seguramente más grandes) no tienen la presión de moverse tan rápido como las empresas más nuevas con menor recorrido (y más pequeñas). Las empresas con mayor recorrido podrían tener ya un producto consolidado, lo que les daría flexibilidad con respecto a la velocidad en el envío de software.

Cuando la información fluye sin problemas, las tareas salen adelante. Descubrimos que unos niveles más altos de intercambio de información se asocian a un mayor rendimiento del envío de software y operativo. Cuando es fácil acceder a la información y existen pocos silos de conocimiento, las personas pueden invertir tiempo en tareas importantes en lugar de ir buscando la información necesaria para hacer esas tareas.

Por último, las modalidades de trabajo flexibles, en las que los empleados pueden determinar cuándo, dónde y cómo trabajan, tienen un efecto beneficioso en todas las métricas de rendimiento. Especialmente, en el rendimiento del envío de software. Incluso si las empresas endurecen sus políticas de teletrabajo, es posible que sea beneficioso permitir una cierta flexibilidad a los empleados.

Aspecto de la cultura	Efecto en el rendimiento del equipo	Efecto en el rendimiento empresarial	Efecto en el rendimiento del envío de software	Efecto en el rendimiento operativo
Cultura empresarial de Westrum	Incremento sustancial	Incremento sustancial	Incremento sustancial	Incremento sustancial
Estabilidad empresarial	Incremento pequeño	Incremento sustancial	Reducción pequeña	Ningún efecto
Seguridad laboral	Incremento pequeño	Ningún efecto	Incremento pequeño	Incremento pequeño
Flexibilidad	Incremento pequeño	Incremento pequeño	Incremento sustancial	Incremento pequeño
Intercambio de conocimientos	Incremento pequeño	Reducción pequeña	Incremento sustancial	Incremento sustancial
Enfoque centrado en el usuario	Incremento sustancial	Incremento sustancial	Incremento pequeño	Incremento sustancial
Distribución del trabajo	Incremento sustancial	Incremento sustancial	Reducción sustancial	Ningún efecto



Una cultura sana mejora las competencias técnicas

Nuestras conclusiones indican que una buena cultura ayuda a mejorar la implementación de competencias técnicas. Creemos que entre la cultura y las competencias técnicas existe una relación recíproca: la cultura surge de las prácticas, y las prácticas surgen de la cultura.

La cultura es amplia y cuesta definirla, mientras que las competencias técnicas suelen estar acotadas y bien definidas. Esto repercute en la manera en que los individuos de una empresa pueden ayudar a impulsar el cambio.

Por ejemplo, los líderes pueden crear estructuras de incentivos que promuevan una cultura generativa.

Tanto ellos como los colaboradores individuales pueden insistir en el uso de un enfoque centrado en el usuario para el desarrollo de software. Los colaboradores individuales pueden ayudar a orientar la implementación de las competencias técnicas que mejoran el rendimiento: desarrollo basado en troncales, integración continua, prácticas de fiabilidad y arquitectura con bajo acoplamiento. Implementar estas competencias técnicas no es sencillo y, para hacerlo bien, es necesario que las personas trabajen codo con codo, tengan una mentalidad abierta, se apoyen mutuamente y aprendan unas de otras. Todo esto da lugar a una cultura sana. Estos equipos pasan a servir de ejemplo para otros equipos de la empresa, que podrían sentirse más capacitados para impulsar el cambio utilizando los medios a su alcance.

Aspecto de la cultura	Efecto en el desarrollo basado en troncales	Efecto en las prácticas de fiabilidad	Efecto en la integración continua	Efecto en la entrega continua	Efecto en la arquitectura con bajo acoplamiento
Cultura empresarial de Westrum					
Estabilidad empresarial					
Seguridad laboral					
Flexibilidad					
Intercambio de conocimientos					
Enfoque centrado en el usuario					
Distribución del trabajo					

Una cultura sana mejora el bienestar del empleado

Una cultura sana contribuye a incrementar los niveles de bienestar de los empleados al disminuir el desgaste y aumentar la satisfacción laboral y la productividad. El bienestar de los empleados no solo es algo deseable, es fundamental para el estado general y el éxito de una empresa.

¿Qué sucede cuando las empresas *no* invierten en una cultura mejor? La posibilidad de sufrir desgaste

aumenta y la satisfacción con el trabajo disminuye. Los empleados se vuelven escépticos y su productividad baja. Su salud física y mental también se ve afectada de forma negativa.^{2,3} El desgaste se cronifica; no basta con tomarse unos días libres para recuperarse. El desgaste también aumenta la rotación: los empleados se marchan en busca de entornos laborales más sanos.⁴ Por ello, reducir el desgaste requiere hacer cambios en las empresas para que se aborden sus causas.

Aspecto de la cultura	Efecto en el desgaste*	Efecto en la satisfacción laboral	Efecto en la productividad
Cultura empresarial de Westrum	 Reducción sustancial	 Incremento sustancial	 Incremento sustancial
Estabilidad empresarial	 Reducción sustancial	 Incremento sustancial	 Incremento pequeño
Seguridad laboral	 Reducción sustancial	 Incremento pequeño	 Incremento pequeño
Flexibilidad	 Reducción pequeña	 Incremento pequeño	 Incremento pequeño
Intercambio de conocimientos	 Reducción sustancial	 Incremento pequeño	 Incremento pequeño
Enfoque centrado en el usuario	 Reducción pequeña	 Incremento sustancial	 Incremento sustancial
Distribución del trabajo	 Ningún efecto	 Incremento pequeño	 Incremento pequeño

* Fíjate en cómo se invierte el esquema de colores en el caso del desgaste porque reducirlo es algo positivo.

² Adam Bayes, Gabriela Tavella & Gordon Parker (2021) "The biology of burnout: Causes and consequences", The World Journal of Biological Psychiatry, 22:9, 686–698. DOI: 10.1080/15622975.2021.1907713. <https://doi.org/10.1080/15622975.2021.1907713>

³ Maslach C, Leiter MP. "Understanding the burnout experience: recent research and its implications for psychiatry". World Psychiatry, junio del 2016, 15(2), 103–11. DOI: 10.1002/wps.20311. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4911781/>. PMID: 27265691; PMCID: PMC4911781.

⁴ L.A. Kelly et al. "Impact of nurse burnout on organizational and position turnover". Nursing Outlook, enero del 2021, 96–102, enero-febrero del 2021, 96–102. DOI: doi.org/10.1016/j.outlook.2020.06.008

Capítulo 8

Cómo, cuándo y por qué importa ser quien eres

Datos clave

Quien eres importa: hemos descubierto que los resultados de algunos grupos de encuestados son diferentes a los de otros, por ejemplo en términos de un desgaste mayor o una productividad menor. También hemos identificado prácticas específicas que puedes implementar para mitigar algunos de estos resultados negativos.

Introducción

Una idea general protagonizó el análisis del 2022: la forma en la que se hace el trabajo podría favorecer el bienestar de algunos empleados, pero no de todos.

En el 2022 descubrimos que las personas que se identificaban como pertenecientes a grupos infrarrepresentados tenían unos niveles de desgaste más altos.¹ En este capítulo veremos que este resultado se repite, y empezaremos a analizar el motivo por el que es más probable que los grupos infrarrepresentados experimenten desgaste y qué factores pueden ayudar a prevenirllo.

Además, la inestabilidad que se ha apoderado de numerosos sectores ha suscitado dudas en torno a nuevas contrataciones. A las empresas les preocupa que los nuevos empleados tarden mucho tiempo en ser productivos. Están buscando nuevas formas de ayudar a los nuevos empleados a ponerse al día más rápido. También profundizaremos en este tema.

¿Qué hemos descubierto y qué significa?

Algunas personas experimentan un desgaste mayor que otras.

El año pasado descubrimos que los encuestados que se identificaban como mujeres o que autodeterminaban su género y los que se identificaban como pertenecientes a grupos infrarrepresentados afirmaban tener un desgaste mayor que los encuestados que se identificaban como hombres no pertenecientes a grupos infrarrepresentados. Estas conclusiones están en línea con un conjunto de investigaciones previas que indican que las personas infrarrepresentadas experimentan un mayor grado de desgaste² y estrés relacionado con el trabajo³ que sus compañeros que sí se sienten representados.

Por ello, nos interesaba saber si este año también había diferencias en nuestros datos sobre el desgaste. Y las encontramos. Los encuestados que se identificaban como mujeres o que autodeterminaban su género señalaron experimentar unos niveles de desgaste un 6 % mayores que los encuestados que se identificaban como hombres. Los encuestados que se identificaban como pertenecientes a grupos infrarrepresentados registraron unos niveles de desgaste un 24 % mayores que los encuestados que no se identificaban como pertenecientes a estos grupos.

¹ Informe Accelerate State of DevOps 2022. <https://dora.dev/research/2022/dora-report/>

² Sigalit Ronen y Ayala Malach Pines, "Gender Differences in Engineers' Burnout", Equal Opportunities International, 7 de noviembre del 2008, <https://www.emerald.com/insight/content/doi/10.1108/02610150810916749/full.html>

³ Dalessandro C; Lovell A; Tanner OC, "Race, Marginalization, and Perceptions of Stress Among Workers Worldwide Post-2020". Sociological Inquiry, 3 de agosto del 2023, <https://onlinelibrary.wiley.com/doi/10.1111/soin.12505>

Algunos tipos de trabajo conducen a un desgaste mayor

Hemos descubierto que aspectos relacionados con el lugar de trabajo que podría parecer que tienen efectos neutros o beneficiosos, como una documentación de calidad o un equipo estable, no disminuyen el desgaste para todos los individuos (consulta el [Capítulo 4 - La documentación es fundamental](#)). No sabemos si esto podría deberse a que hay tareas que, aunque benefician a la empresa, contribuyen al desgaste de algunos individuos.⁴

Para entender la experiencia de desgaste de los encuestados, les hemos preguntado sobre el trabajo que realizan. Lo hemos medido según dos variables:

- Tareas específicas, como programación, reuniones o asistencia a compañeros del equipo.
- Características del trabajo, como tareas imprevistas, su visibilidad o la cantidad de labor manual.

Las características del trabajo son importantes, porque la misma tarea podría experimentarse de diferente manera por diferentes personas o en diferentes momentos. Por ejemplo, algunas revisiones de código pueden ser una labor manual imprevista, mientras que otras pueden tener un nivel de visibilidad altísimo en el equipo y demostrar liderazgo y conocimientos técnicos.⁵

Los encuestados que se identificaban como infrarrepresentados señalaron hacer un 24 % más de trabajo repetitivo (labor manual) que aquellos que no se identificaban como infrarrepresentados. Los encuestados que se identificaban como mujeres o que autodeterminaban su género señalaron hacer un 40 % más de trabajo repetitivo que los encuestados que se identificaban como hombres. Ambos grupos de encuestados también afirmaban hacer más tareas imprevistas y trabajo que no tenía tanta visibilidad para los demás ni se relacionaba directamente con sus habilidades profesionales. Estas conclusiones explican parcialmente el desgaste manifestado por estos grupos.

⁴ Linda Babcock, Brenda Peyser, Lise Vesterlund y Laurie Weingart. *The No Club* (New York: Simon and Schuster, 2022), 17.

⁵ Murphy-Hill, E. et al. "Systemic Gender Inequities in Who Reviews Code", Computer Supported Cooperative Work (2023), <https://research.google/pubs/pub52204>

⁶ Babcock et al., *The No Club*, 17.

Tareas que no sirven para ascender

En su libro *The No Club*, Babcock et al. describen un tipo de trabajo específico: "Una tarea que no sirve para ascender le importa a tu empresa, pero no te ayudará a avanzar en tu carrera".⁶

Aquí la palabra *avanzar* se usa en sentido amplio y se refiere, por ejemplo, a una mayor remuneración o a un aumento de las posibilidades de acceder a otros empleos.

Se ha demostrado que las mujeres hacen más trabajo de este tipo. Babcock et al. describen los motivos de esta distribución desigual del trabajo. Por ejemplo, es más probable que este tipo de tareas se les pida hacer a las mujeres, y es más probable que digan que sí, ya que decir que no conlleva un coste social.

También describen las consecuencias de la distribución desigual de estas tareas. Algunas mujeres, por ejemplo:

- Observan que afecta de forma negativa a su carrera profesional o sus beneficios.
- Hacen más horas para tener un volumen adecuado de trabajo que resulte relevante para su carrera.

Los procesos formales de distribución del trabajo disminuyen el desgaste en algunos encuestados

Hemos preguntado a los encuestados si siguen procesos formales para distribuir el trabajo de manera equitativa en sus empresas. Lo llamamos **distribución del trabajo** y esperábamos que mitigara el desgaste que experimentan algunos de los encuestados.

Lo que descubrimos fue que la distribución del trabajo reducía el desgaste en el caso de los encuestados que se identificaban como hombres, los que se identificaban como mujeres y los que autodeterminaban su género. Si el nivel de distribución del trabajo era alto, la diferencia del desgaste entre géneros desaparecía.

Nos sorprendió descubrir que la distribución del trabajo no influía nada en el nivel de desgaste experimentado por los encuestados que se identificaban como infrarrepresentados. Este resultado plantea más preguntas: ¿Los procesos formales para distribuir el trabajo de manera equitativa siguen derivando en una distribución desigual del trabajo? ¿El "trabajo igualitario" tiene en cuenta las características de las tareas, como las interrupciones o la visibilidad? Y ¿cómo mitigamos otros factores que contribuyen al desgaste, además de las tareas en el trabajo, que podrían ser más importantes para este grupo?



Fomentar una cultura de pertenencia

Autor: Dr. Jeffrey Winer, psicólogo asistencial en el Boston Children's Hospital, profesor auxiliar en la Facultad de Medicina de Harvard

Un resultado clave y un poco de contexto

Un resultado clave de este informe es que las personas que se definen como pertenecientes a grupos infrarrepresentados experimentan un desgaste mucho mayor que sus compañeros. En el informe se han estudiado posibles motivos. En esta sección, vincularemos estas conclusiones con una investigación más amplia sobre pertenencia y estrategias de prácticas empresariales asociadas.

Identificarse como infrarrepresentado en un grupo muestra una vulnerabilidad a la "incertidumbre de pertenencia",⁷ un fenómeno psicológico consolidado (consulta "Understanding and Overcoming Belonging Uncertainty", del Dr. Geoffrey Cohen).⁸ Esa incertidumbre (por ejemplo, "¿Pertenezco a este lugar?", "¿Alguien como yo puede triunfar aquí?") se refuerza o redefine por las experiencias continuas de las personas y las interpretaciones de esas experiencias. Estos procesos consolidados asociados a la incertidumbre de pertenencia pueden servir para contextualizar lo que hemos descubierto en este informe sobre las personas que se identifican como infrarrepresentadas, que registran unos niveles de desgaste más altos.

¿Qué pueden hacer las empresas?

Es importante recordar que la diversidad, la inclusión, la igualdad y la pertenencia significan cosas distintas y que, para lograrlas, requieren estrategias distintas, interconectadas y continuadas. Alcanzar un sentido de pertenencia requiere compromisos auténticos y con continuación.

Si alguien tiene problemas en su empresa, lo primero que habría que preguntarse no debería ser: "¿Qué le pasa a esta persona?", sino "¿Por qué tendría sentido que alguien se sintiera así y qué elementos estructurales de nuestra empresa promueven ese sentimiento (por ejemplo, qué elementos mantienen esa sensación o la empeoran)?".

Cuando se identifica el problema, los cambios deben hacerse a nivel de empresa, al mismo tiempo que se da apoyo a las personas individualmente, mediante un enfoque integrador. Apoyar a los individuos para que influyan en los sistemas que rigen una empresa permitirá integrar cambios en el sistema y que perduren aunque varíen los actores individuales. Adoptar esta mentalidad sistemática y de continuidad permitirá integrar los cambios en la institución de manera que estos perduren aunque varíen los actores individuales. Esta calidad generativa es lo que permite a las empresas esforzarse por lograr un sentido de pertenencia. Esforzarse es vital. La pertenencia se construye a través de la experiencia y las acciones continuadas; nunca termina, y por eso es imprescindible para la salud y la productividad del lugar de trabajo.

Existen una serie de herramientas para apoyar a las empresas en esta tarea. Por ejemplo, el informe sobre la soledad del 2023 de Surgeon General, administración que supervisa la salud pública de EE. UU., señala que la conexión social y la pertenencia son los principales antídotos contra la soledad y el desgaste.⁹

⁷ Walton GM; Cohen GL; "A Brief Social-Belonging Intervention Improves Academic and Health Outcomes of Minority Students", Science (Nueva York, NY), fecha de consulta: 20 de septiembre del 2023, <https://pubmed.ncbi.nlm.nih.gov/21415354/>

⁸ <https://behavioralscientist.org/understanding-and-overcoming-belonging-uncertainty/>

⁹ Oficina del Subsecretario de Salud (OASH), "New Surgeon General Advisory Raises Alarm about the Devastating Impact of the Epidemic of Loneliness and Isolation in the United States", HHS.gov, 3 de mayo del 2023, <https://www.hhs.gov/about/news/2023/05/03/new-surgeon-general-advisory-raises-alarm-about-devastating-impact-epidemic-loneliness-isolation-united-states.html>

Las nuevas contrataciones tienen problemas de productividad

Las nuevas contrataciones (menos de 1 año de experiencia en el equipo) tienen una puntuación de productividad un 8 % más baja que sus compañeros con más experiencia (más de 1 año de experiencia), lo que posiblemente es previsible. Empezar en un equipo nuevo es todo un reto y, aunque tengas experiencia en el puesto, la cantidad de conocimientos específicos del equipo que se necesitan para ponerse en marcha puede ser abrumadora. Además, formar parte de un equipo conlleva algo más que tener unas habilidades y unos conocimientos. De manera anecdótica, también hay un componente social que es imprescindible para la

productividad. Lleva un tiempo desarrollar aspectos como la pertenencia y la seguridad psicológica, así como sentir que aportas algo al equipo.

¿Hay que algo que pueda ayudar a las nuevas contrataciones a acelerar el proceso de esos aspectos?

Planteamos la hipótesis de que las empresas podrían ayudar a las nuevas incorporaciones con tres acciones:

- Facilitar documentación de alta calidad.
- Incorporar inteligencia artificial en los flujos de trabajo, lo que en otro estudio ha demostrado ser más útil para los trabajadores sin experiencia que para los experimentados.
- Trabajar en persona codo con codo, que podría ser especialmente beneficioso durante la fase de formación inicial.



Los resultados de este año sugieren que una documentación de alta calidad influye sustancialmente en la productividad (consulta el [Capítulo 4 - La documentación es fundamental](#)) y que los efectos positivos de la IA en la productividad de un individuo son insignificantes (consulta el [Capítulo 3 - Las competencias técnicas conducen a un mayor rendimiento](#)). No hay nada que nos haga pensar que estos efectos no se extiendan también a las nuevas contrataciones. Si nos fijamos en los datos, vemos que este tipo de prácticas ayuda a las nuevas contrataciones, pero ni más ni menos que al resto de personas. Dicho de otra manera, las nuevas contrataciones no se benefician de una forma especial de estas prácticas.

Si estás buscando maneras de ayudar a las nuevas contrataciones y al resto de empleados, facilitar una documentación de alta calidad es un primer paso excelente, ya que tiene un efecto continuado y claro en la productividad. Cabe señalar que las nuevas contrataciones en equipos que disponen de documentación bien elaborada (1 desviación estándar por encima de la media) son un 130 % más productivas que las nuevas contrataciones en equipos con una documentación elaborada de manera deficiente (1 desviación estándar por debajo de la media).

No nos extenderemos en el debate sobre la vuelta a la oficina para no echar más leña al fuego. Hablamos de la importancia de la flexibilidad en el [Capítulo 7 - Nada de esto es posible si no se invierte en la cultura](#). Además, nuestros datos no son experimentales y, aunque tratamos de controlar los factores que podrían condicionar los resultados, las ventajas de la modalidad de trabajo son un asunto complejo y enrevesado desde el punto de vista sociológico. Por eso, es difícil sacar conclusiones definitivas (conviene recordarlo al leer investigaciones o artículos de opinión sobre el tema). Lo que sí indican claramente nuestros datos es que la flexibilidad tiene un efecto positivo en la productividad. Pero no queda tan claro si el lugar de trabajo también tiene ese efecto.

Lo mismo ocurre con las nuevas contrataciones. No hemos observado ninguna evidencia de que trabajar codo con codo de manera presencial suponga una ventaja especial para las nuevas contrataciones. Si tu objetivo es optimizarlo todo para aumentar la productividad, parece que dar flexibilidad a las nuevas contrataciones con respecto a cómo, dónde y cuándo trabajan puede ser una apuesta más segura que obligarlos a ir a la oficina. Está claro que las empresas no buscan la optimización solo por mejorar la productividad (ni deberían). También definimos la productividad como tareas que aportan valor, no simples resultados (y no hablamos de líneas de código), así como trabajo que no causa desgaste ni labor manual.



Epílogo

Reflexiones finales

Gracias por participar en la investigación de este año y leer este informe. Siempre estamos buscando optimizar la forma de estudiar las conexiones entre la manera de trabajar de los equipos y los resultados que son capaces de alcanzar.

La conclusión más importante que nos llevamos de nuestro programa de investigación estos años es que los equipos que adoptan una mentalidad y una práctica de mejora continua son capaces de lograr los mejores resultados.

Las competencias que hemos estudiado se pueden usar como factores que ajustar para impulsar resultados. Algunos de ellos están al alcance de los empleados, mientras que a otros solo se puede acceder a través del esfuerzo coordinado de toda la empresa. Identifica los factores que necesitas ajustar en tu empresa y, después, invierte en ellos.

El trabajo de mejora nunca termina, pero puede traer el éxito a largo plazo para los individuos, los equipos y las empresas. Los responsables y los profesionales comparten la responsabilidad de impulsar este trabajo de mejora.

¿Cómo vas a poner en práctica esta investigación?

Identifica estas conclusiones en el contexto de tu empresa, tus equipos y los servicios que ofreces a tus clientes.

Únete a la comunidad DORA para compartir tus experiencias, aprender de los demás e inspirarte en los compañeros con los que recorres el camino hacia la mejora continua: <https://dora.community>.

Agradecimientos

Este informe recibe todos los años el apoyo de una gran familia de apasionados colaboradores de todas partes del mundo. En cada fase de su elaboración, desde el diseño de las preguntas de la encuesta y la localización hasta el análisis, la redacción, la revisión y la composición tipográfica, han participado compañeros que han ayudado a completar una tarea de semejante magnitud. Los autores quieren dar las gracias a todas estas personas por sus contribuciones, asesoramiento y camaradería.

Colaboradores

Equipo principal

James Brookbank

Kim Castillo

Derek DeBellis

Nathen Harvey

Michelle Irvine

Amanda Lewis

Eric Maxwell

Steve McGhee

Dave Stanke

Kevin Storer

Daniella Villalba

Brenna Washington

Editores

Mandy Grover

Jay Hauser

Stan McKenzie

Anna Eames Mikkawi

Mike Pope

Tabitha Smith

Olinda Turner

Localización de la encuesta

Daniel Amadei

Kuma Arakawa

William Bartlett

Antonio Guzmán

Shogo Hamada

Yuki Iwanari

Vincent Jobard

Gustavo Lapa

Mauricio Meléndez

Jeremie Patonnier

Miguel Reyes

Pedro Sousa

Laurent Tardif

Kimmy Wu

Vinicius Xavier

Yoshi Yamaguchi

Asesores y expertos en la materia

Jared Bhatti

Lisa Crispin

Rob Edwards

Dave Farley

Steve Fenton

Dr. Nicole Forsgren

Aaron Gillies

Denali Lumma

Emerson Murphy-Hill

Harini Sampath

Robin Savinar

Dustin Smith

Jess Tsimeris

Dra. Laurie Weingart

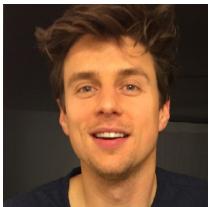
Betsalel (Saul) Williamson

Dr. Jeffrey Winer

Patrocinadores



Autores



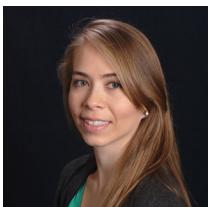
Derek DeBellis

Derek es investigador de experiencia de usuario cuantitativa en Google y jefe de la investigación de DORA. Lleva a cabo investigaciones por medio de encuestas, analiza registros y estudia formas de medir conceptos que demuestren que un producto o una función aporta un gran valor a las personas. Derek ha publicado artículos de investigación sobre la interacción entre los humanos y la inteligencia artificial, el impacto del COVID-19 en el abandono del tabaquismo, el diseño para abordar los errores de PLN, el papel de la experiencia de usuario en los debates sobre la privacidad, la cultura de equipo y la relación de la IA con el bienestar y la productividad de los empleados. En su actual investigación curricular analiza formas de simular la propagación de las creencias y el poder.



Amanda Lewis

Amanda Lewis es jefa de desarrollo de DORA.community e ingeniera de relaciones con desarrolladores en el equipo de DORA Advocacy de Google Cloud. Ha dedicado su carrera profesional a crear conexiones entre desarrolladores, operadores, responsables de producto, gestores de proyectos y líderes. Ha trabajado en equipos que desarrollaban plataformas de comercio electrónico, sistemas de gestión de contenido y herramientas de observabilidad, y ha prestado apoyo a desarrolladores. Estas conexiones y conversaciones mejoran la satisfacción de los clientes y los resultados de la empresa. Aporta su experiencia y empatía a la hora de ayudar a los equipos a entender e implementar el envío de software y las prácticas de fiabilidad.



Daniella Villalba

Daniella Villalba es investigadora de experiencia de usuario en Google. Utiliza la investigación por medio de encuestas para comprender los factores que mejoran la satisfacción y la productividad de los desarrolladores. Antes de trabajar en Google, Daniella estudió las ventajas de aprender meditación y los factores psicosociales que afectan a las experiencias de los estudiantes universitarios. Tiene un doctorado en Psicología Experimental por la Universidad Internacional de Florida.



Dave Farley

Dave Farley es el director ejecutivo y fundador de Continuous Delivery Ltd, autor de *Modern Software Engineering* y coautor del exitoso libro *Continuous Delivery*. Es uno de los autores del *Reactive Manifesto* y ganador del premio Duke por su proyecto de código abierto "LMAX Disruptor". Dave es precursor de la entrega continua, formador de opinión y experto en CD, DevOps, desarrollo guiado por pruebas y diseño de software. Cuenta con una dilatada experiencia en formación a equipos de gran rendimiento, configuración de empresas para el éxito y creación de software excepcional. Dave está comprometido a compartir su experiencia y sus métodos con desarrolladores de software de todo el mundo y ayudarles a mejorar el diseño, la calidad y la fiabilidad de su software. Comparte su experiencia a través de sus servicios de consultoría,¹ canal de YouTube² y cursos de formación.



Eric Maxwell

Eric Maxwell dirige las prácticas de transformación de DevOps en Google, donde asesora a las empresas más importantes del mundo sobre cómo mejorar al ofrecer valor de forma más rápida. Trabajó como ingeniero durante la primera mitad de su carrera, automatizando de todo y abogando por la empatía hacia otros profesionales. Colaboró en la creación del Cloud Application Modernization Program (CAMP) de Google y es miembro del equipo de DORA. Antes de trabajar en Google, Eric estuvo cocinando maravillas con otros tíos ingeniosos en Chef Software.



James Brookbank

James Brookbank es arquitecto de soluciones en la nube de Google. Los arquitectos de soluciones ayudan a los clientes de Google Cloud a resolver problemas técnicos complejos y les ofrecen asesoramiento profesional de arquitectura. Antes de unirse a Google, James trabajó en varias empresas grandes de infraestructura de TI y servicios financieros.



Dr. Jeffrey Winer

Jeffrey P. Winer tiene un doctorado y es psicólogo asistencial, consultor en sistemas de salud del comportamiento y desarrollador de tratamientos psicosociales en el Boston Children's Hospital Trauma and Community Resilience Center (TCRC), además de profesor auxiliar en la Facultad de Medicina de Harvard. Junto a sus compañeros del TCRC, su trabajo se centra principalmente en crear, probar, difundir e implementar intervenciones psicosociales que tienen en cuenta la cultura y los problemas de salud mental para jóvenes y familias de refugiados e inmigrantes. Es coautor del libro *Mental Health Practice with Immigrant and Refugee Youth: A Socioecological Framework*³ y ha trabajado como consultor en programas en Estados Unidos y Canadá. Las herramientas de prevención e intervención psicosociales que ha ayudado a desarrollar o adaptar se usan actualmente en todo el mundo. Para obtener más información, visita <http://www.drjeffwiner.com>.

¹ <https://continuous-delivery.co.uk/engineering-for-software>

² <https://www.youtube.com/@ContinuousDelivery>

³ <https://www.apa.org/pubs/books/4317536>



Kevin Storer

Kevin M. Storer es investigador de experiencia de usuario en Google, donde lidera la investigación dirigida a entender cómo interactúan los equipos de desarrollo de software con las herramientas de DevOps y a través de ellas. Antes de llegar a Google, Kevin obtuvo un doctorado en Informática por la Universidad de California en Irvine. Kevin es autor de publicaciones de gran impacto tanto en el sector público como en el privado sobre programación centrada en el ser humano, experiencia del desarrollador, comportamiento de la información, accesibilidad y computación ubicua.



Kim Castillo

Kim Castillo es directora de programas de experiencia de usuario en Google y dirige los esfuerzos multifuncionales detrás de DORA, desde supervisar sus operaciones de investigación hasta la publicación del informe. Kim también trabaja en la investigación de la experiencia de usuario de Duet AI en Google Cloud. Antes de Google, ocupaba puestos relacionados con el envío de software, donde trabajaba en la gestión de programas técnicos y el coaching hacia la agilidad. Los orígenes de Kim están en la investigación psicosocial, que dedicó a temas como las ejecuciones extrajudiciales, el desarrollo urbano de zonas pobres y el trauma y la resiliencia de la comunidad en su país de origen, Filipinas. DORA combina varias pasiones de Kim: prácticas de envío de software, seguridad psicológica y empresas "teal".



Michelle Irvine

Michelle Irvine es redactora técnica en Google y dirige la investigación sobre el impacto y la elaboración de documentación técnica. Antes de llegar a Google, trabajó en el ámbito de la edición para centros educativos y como redactora técnica de software de simulación física. Michelle se licenció en Física y tiene un máster en Retórica y Diseño de Comunicación por la Universidad de Waterloo.



Nathen Harvey

Nathen Harvey dirige el equipo de DORA Advocacy como director de ingeniería de relaciones con desarrolladores de Google Cloud. Ha tenido el privilegio de trabajar con algunos de los mejores equipos y comunidades de software libre, ayudándoles a aplicar los principios y las prácticas de DevOps y SRE. Ha sido coautor del informe *Accelerate State of DevOps* en los últimos tres años. Nathen también correvisó y contribuyó a la publicación *97 Things Every Cloud Engineer Should Know*.



Steve McGhee

Steve McGhee es un ingeniero de fiabilidad que ayuda a los equipos a entender cómo crear y gestionar de la mejor manera servicios fiables de primer nivel. Antes, trabajó más de 10 años como ingeniero de fiabilidad del sitio en Google, donde aprendió a escalar sistemas globales en la Búsqueda, YouTube, Android y Google Cloud. Gestionaba varios equipos de ingeniería en California, Japón y el Reino Unido. Steve también trabajó un tiempo con una empresa de California a la que ayudó a pasarse a la nube.

Metodología

Este capítulo describe cómo partimos de un conjunto de ideas iniciales hasta llegar al informe que tienes ante ti (con ideas un poco más completas). Esperamos que responda a muchas de tus preguntas sobre cómo se elabora y que te sirva de modelo para iniciar tu propia investigación.

Paso 1: generar un conjunto de resultados que consideramos importantes para empresas tecnológicas con un alto rendimiento

Esto es fundamental. La base de nuestro programa es ayudar a orientar a las personas hacia los fines que les interesan. Si no sabemos a dónde quieren llegar las personas, las empresas o los equipos, entonces ya empezamos mal. ¿Cómo lo averiguamos? Combinamos investigación cualitativa (es decir, preguntar a las personas qué quieren alcanzar ellas, sus equipos y sus empresas), encuestas, interacción con otros miembros de la comunidad y muchos talleres. Obtenemos sistemáticamente resultados como los siguientes:

- **Rendimiento empresarial.** La empresa no solo debe generar ingresos, sino también aportar valor al cliente y a la comunidad.
- **Rendimiento del equipo.** La capacidad de un equipo de aplicaciones o servicios para crear valor, innovar y colaborar.
- **Bienestar de los empleados.** Las estrategias que adopta una empresa o un equipo deben beneficiar a los empleados: reducir el desgaste, promover una experiencia laboral satisfactoria y aumentar la capacidad que tienen las personas de producir resultados valiosos (es decir, la productividad).

También escuchamos hablar de objetivos como los siguientes:

- **Rendimiento del envío de software.** Los equipos despliegan software de forma rápida y eficaz.
- **Rendimiento operativo.** El software enviado ofrece una experiencia fiable para el usuario.



Paso 2: plantear hipótesis sobre cómo, cuándo y por qué se alcanzan estos resultados

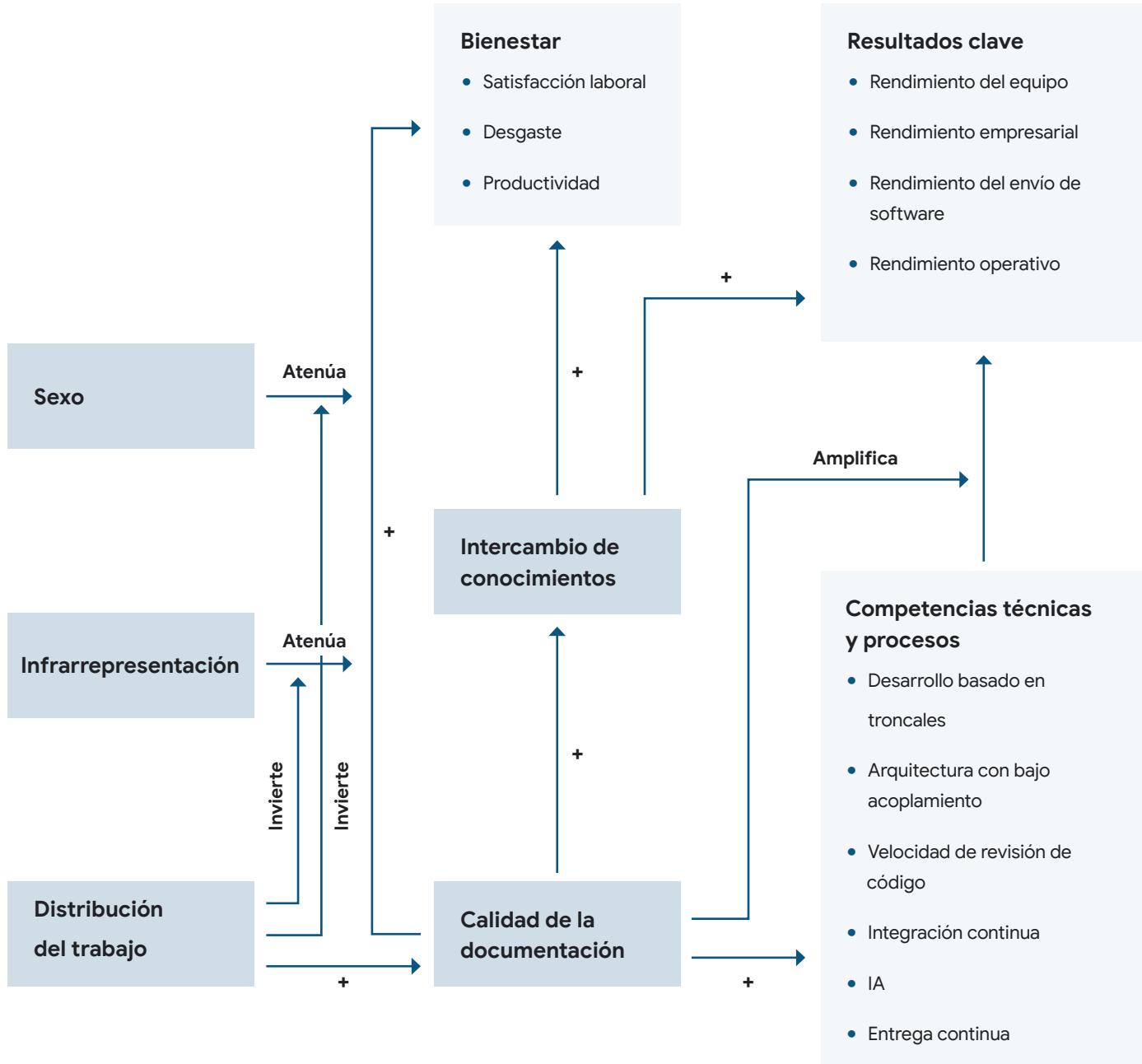
Tenemos en cuenta los resultados del paso 1 para formular hipótesis sobre cómo podrían alcanzarlos los equipos. Esto implica buscar factores que tengan un efecto fiable en el resultado. Queremos llegar a la conclusión "si todo se mantiene igual, X tiene un efecto sobre Y". Esta información sirve para ayudar a los profesionales a tomar decisiones basadas en datos sobre el tipo de cambios que deben probar.

No solo queremos entender qué procedimientos generan un efecto, sino también analizar en qué *condiciones* estos procedimientos producen un mayor o menor efecto. Esto equivale a preguntar "cuándo" y "para quién". Por ejemplo, se ha demostrado que, de media, la calidad de la documentación reduce drásticamente el desgaste. Sin embargo, los encuestados que se identificaban como infrarrepresentados indicaban lo contrario: la calidad de la documentación *aumenta* el desgaste. Resulta esencial entender las condiciones en las que se producen estos efectos, porque es muy poco habitual que un equipo o una persona se sitúe en la media.

Asimismo, formulamos hipótesis sobre *mecanismos* que explican *por qué* o *cómo*. Este año planteamos la hipótesis de que las personas que se identificaban como infrarrepresentadas experimentaban un desgaste mayor, en función de los resultados que observamos el año pasado y de una amplia bibliografía sobre el tema. Lo primero que nos preguntamos fue: "¿Por qué pasa esto?". Para tratar de responder a esta pregunta, planteamos hipótesis acerca de posibles mecanismos a evaluar. Por ejemplo, las personas que se identifican como infrarrepresentadas podrían experimentar un desgaste mayor *porque* asumen (o se les asignan) más tareas tediosas.



Estas hipótesis se desarrollan después para que podamos crear la encuesta y realizar los análisis. A continuación puedes ver un ejemplo de un modelo hipotético del capítulo de documentación, que se completa con efectos, condiciones y mecanismo:

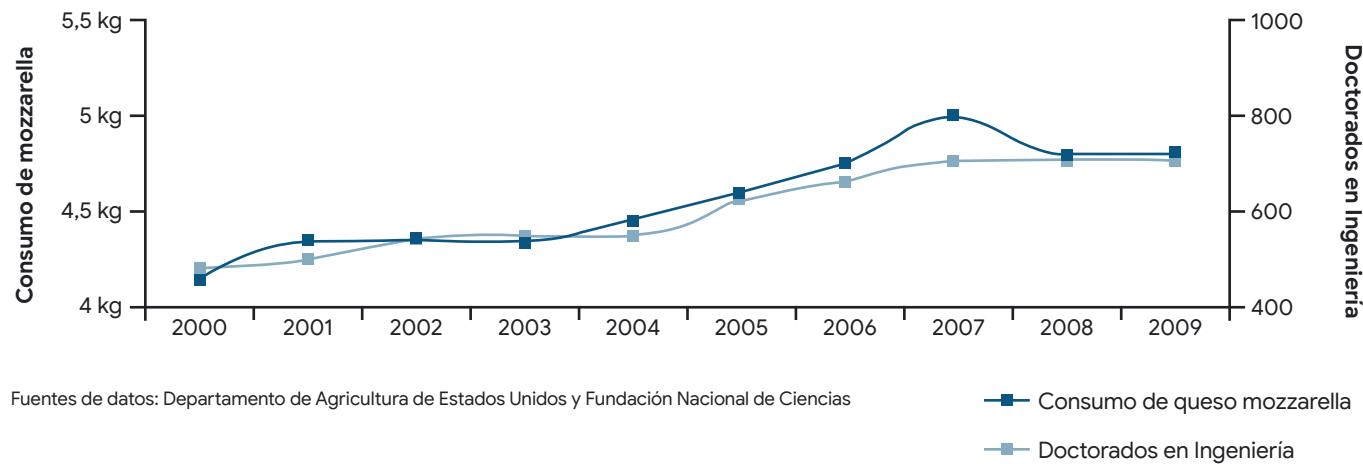


Paso 3: formular hipótesis sobre posibles confusiones

Si alguna vez has tratado temas relacionados con datos, probablemente te hayas encontrado con una correlación espuria. Quizá te suene un sitio web que muestra numerosas correlaciones espurias,¹ como el siguiente ejemplo:

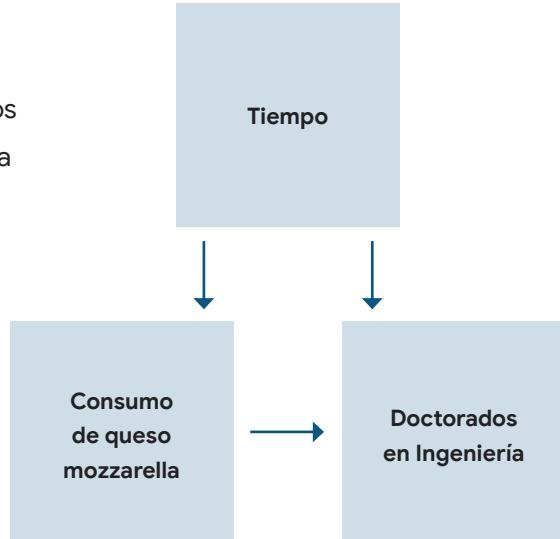
El consumo de queso mozzarella per cápita está relacionado con la obtención de títulos de doctorado en Ingeniería civil

Correlación: 95,86 % ($r=0,958648$)



Es muy poco probable que exista una relación de causalidad entre los doctorados en Ingeniería y el consumo de queso mozzarella. Sin embargo, detrás de esta relación se esconde un elemento confuso: el tiempo. Si la tendencia del consumo de mozzarella y del número de doctorados en Ingeniería es positiva en el mismo periodo, es probable que tengan una correlación positiva.

Incluir el tiempo en un modelo o eliminar las tendencias de los datos probablemente anulará la relación. Podemos dibujar el modelo de la siguiente manera:



¹ <https://www.tylervigen.com/spurious-correlations>

Si no tenemos en cuenta el tiempo (una tercera variable), los datos podrían mostrar una relación espuria entre el consumo de queso mozzarella y los doctorados en Ingeniería.

Existen herramientas que facilitan la tarea de los investigadores en este sentido, como es el caso de Dagitty (<https://dagitty.net/dags.html>). Esta herramienta nos permite especificar nuestro modelo causal. Para ayudarnos a calcular de forma adecuada el efecto de X sobre Y, indica las implicaciones del modelo, lo que debemos tener en cuenta y lo que *no* deberíamos incluir. Herramientas como Dagitty pueden llevarnos a la conclusión de que la correlación podría no implicar causalidad, pero da entender el modo en el que las personas *perciben* la causalidad.

Es imposible recoger todos los elementos que pueden condicionar los cálculos de los investigadores; solo hay que pensar en encuestas de 50 horas de duración y en la omnisciencia. Aun así, hacemos todo lo que está en nuestra mano para tener en cuenta todo lo que genera sesgos y así poder ofrecerte cálculos precisos de los efectos que tienen diversas actividades, tecnologías y estructuras en los resultados que te interesan. Al fin y al cabo, muchos profesionales quieren saber qué factores influirán en estos resultados clave. Los modelos que no tienen en cuenta los sesgos no conseguirán ofrecer a los profesionales la orientación que necesitan. No queremos poner en un informe que el queso mozzarella aumenta el rendimiento del envío de software. Y cometer ese error sería más fácil de lo que te imaginas.



Paso 4: desarrollar la encuesta

En el desarrollo de la encuesta intervienen tres aspectos: la operacionalización, la experiencia y la localización.

Operacionalización

Queremos medidas que capten de manera adecuada los conceptos que nos interesan, y que lo hagan con fiabilidad. Traducir un concepto abstracto en algo medible es el arte de la operacionalización. Estos indicadores son los ingredientes básicos de cualquier análisis. Si las medidas no nos aportan señales claras, ¿cómo vamos a confiar en el resto del análisis? ¿Cómo medimos un concepto tan escurridizo como, por ejemplo, la productividad? ¿Qué pasa con el desgaste o el rendimiento operativo?

En primer lugar, consultamos la bibliografía para ver si ya existen medidas de éxito. Si podemos usar medidas validadas previamente en nuestra encuesta, construiremos un puente entre la encuesta y toda la bibliografía acumulada en torno a esa cuestión. El uso continuado que hacemos de la *tipología de la cultura empresarial* de Westrum es un ejemplo de reutilización de medidas validadas con anterioridad.

Sin embargo, hay numerosos conceptos que no han sido validados previamente para el espacio en el que realizamos nuestra investigación. En estos casos, hacemos una investigación cualitativa para descubrir cómo entienden las personas el concepto, y buscamos en la bibliografía más filosófica los entresijos del concepto.

Experiencia de la encuesta

Queremos que la encuesta sea comprensible, sencilla, accesible y no más larga de lo necesario. Son unos objetivos difíciles por todas las preguntas que queremos hacer, por la comprensión técnica que se requiere para responderlas y por la variación en la nomenclatura de determinadas prácticas. Hacemos evaluaciones no moderadas a distancia para asegurarnos de que la encuesta funciona por encima de determinados umbrales. Esto requiere numerosas iteraciones.

Localización

Todos los años responden a nuestra encuesta personas de todo el mundo. Este año hemos trabajado en que la encuesta sea más accesible para una audiencia mayor, y la hemos localizado a español, francés, inglés, japonés y portugués. Esta iniciativa liderada por miembros increíbles de la comunidad DORA ha sido posible gracias a un esfuerzo conjunto de todos. Googlers de todo el mundo han colaborado en la tarea, así como un partner en la materia: muchas gracias a Zenika (<https://www.zenika.com>) por la localización al francés. Esperamos poder ampliar el trabajo y convertir la encuesta en una actividad realmente intercultural.

Paso 5: recopilar las respuestas de la encuesta

Utilizamos múltiples canales para buscar encuestados. Estos canales se engloban en dos categorías: *enfoque orgánico* y *enfoque de panel*.

El *enfoque orgánico* sirve para usar todos los medios sociales a nuestra disposición para que las personas se enteren de que hemos hecho una encuesta en la que queremos que participen. Creamos entradas de blog, lanzamos campañas por correo electrónico, publicamos en redes sociales y les pedimos a los miembros de la comunidad que también lo hagan (es decir, muestreo por bola de nieve).

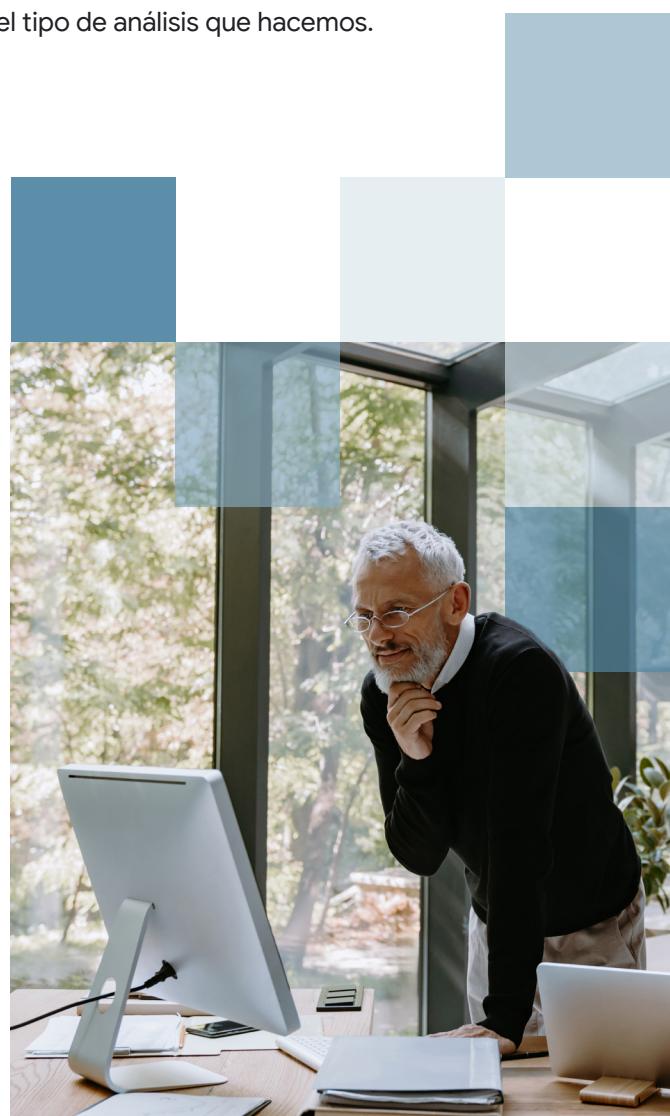
El *enfoque de panel* lo usamos para complementar el canal orgánico. Intentamos buscar a personas tradicionalmente infrarrepresentadas en la comunidad técnica general y tratamos de obtener respuestas adecuadas de determinados sectores y tipos de empresa. En resumen, este enfoque nos permite controlar un poco la captación, algo que contrasta con el enfoque orgánico, donde el control es inexistente. El enfoque de panel también nos permite asegurarnos de que tenemos suficientes encuestados, ya que nunca sabemos si el enfoque orgánico va a conseguir las respuestas necesarias para el tipo de análisis que hacemos.

Paso 6: analizar los datos

El análisis se compone de tres pasos: la limpieza de datos, la validación de mediciones y la evaluación de modelos.

Limpieza de datos

El objetivo de la limpieza de datos es aumentar la relación señal/ruido. Por varios motivos, algunas de las respuestas son ruido. Algunos patrones de respuesta pueden indicar que la persona está distraída, está haciendo la encuesta a toda prisa o no responde de buena fe. Es poco habitual en nuestros datos, pero puede pasar. Las respuestas que dan la impresión de tergiversar las experiencias reales se excluyen de nuestro análisis porque añaden ruido. Al deshacernos del ruido nos enfrentamos al reto de tener la seguridad de que no estamos eliminando la señal, sobre todo de una manera sesgada o que valide nuestras hipótesis. Por ejemplo, si llegamos a la conclusión de que nadie puede tener un nivel alto en un determinado valor y bajo en otro, podríamos estar excluyendo a algún encuestado. De esta forma, los datos se ajustarían más a nuestras creencias y aumentarían las probabilidades de que nuestras hipótesis se cumplieran.



Validación de mediciones

Al comienzo de este informe, hablamos de los conceptos que intentamos medir. Podríamos sumarnos a numerosos y diferentes juegos lingüísticos, pero uno de ellos denomina variable a la medida de un concepto.² Las variables son los ingredientes de los modelos, que a su vez son los elementos incluidos en nuestra investigación. Hay dos formas amplias de analizar la validez de estas medidas: interna y externamente.

Para entender la validez interna de la medida, hay que analizar lo que creemos que indica la presencia de un concepto. Por ejemplo, las personas que mencionan la documentación de calidad podrían ser las que usan la documentación para solucionar sus problemas.

La mayoría de nuestras variables se componen de varios indicadores, porque las ideas que nos interesan parecen ser polifacéticas. Para entender la naturaleza polifacética de una variable, probamos qué tal se combinan los elementos que usamos para representar esa idea. Si se combinan bien (es decir, comparten un nivel alto de varianza comunal), suponemos que hay algo subyacente en ellos, como el concepto de interés.

Pensemos, por ejemplo, en la felicidad. La felicidad es polifacética. Cuando alguien es feliz, esperamos que se sienta de una manera determinada, actúe de una manera determinada y piense de una manera determinada. Damos por hecho que detrás de la felicidad subyace un patrón determinado de sentimientos, pensamientos y acciones. Por eso, cuando la felicidad está presente, esperamos que surjan a la vez determinados tipos de sentimientos, pensamientos y acciones. Entonces plantearíamos preguntas sobre esos sentimientos, pensamientos

y acciones, y usaríamos el análisis factorial confirmatorio para probar si realmente aparecen al mismo tiempo.

Este año hemos utilizado el paquete de R de lavaan para hacer el análisis.³ lavaan ofrece una variedad de estadísticas ajustadas que nos ayudan a entender si los conceptos realmente representan la forma en que las personas responden a las preguntas. Si los indicadores de un concepto no combinan bien, podría ser necesario revisar el concepto o abandonarlo, ya que es evidente que no hemos encontrado una manera fiable de medirlo.

La validez externa de una idea consiste en fijarse en cómo encaja en el mundo. Podríamos esperar que una idea estuviera relacionada de alguna manera con otras ideas. Hay veces que podemos esperar que dos ideas tengan una relación negativa, como la felicidad y la tristeza. Si nuestra medida de felicidad se correlaciona positivamente con la tristeza, quizás debamos cuestionar nuestra medida o nuestra teoría. De forma similar, podríamos esperar que dos ideas tuvieran relaciones positivas, pero no muy sólidas. Es posible que la productividad y la satisfacción laboral tengan una correlación positiva, pero no pensamos que sean idénticas. Si la correlación es demasiado alta, parecería que estuviéramos midiendo lo mismo. Por tanto, llegamos a la conclusión de que nuestras medidas no están lo suficientemente calibradas para recoger las diferencias entre los dos conceptos o la diferencia sobre la que hemos formulado la hipótesis no existe en realidad.

² Moore, Will H. y David A. Siegel. A mathematics course for political and social research. Princeton University Press, 2013.

³ Rosseel, Y. "lavaan: An R Package for Structural Equation Modeling", Journal of Statistical Software, 48(2), 2012. 1–36. <https://doi.org/10.18637/jss.v048.i02>

Evaluación de modelos

En los pasos 2 y 3 hemos creado modelos hipotéticos. Una vez que el paso 6 nos devuelve datos limpios, observamos cómo se ajustan a los datos esos modelos. Este año hemos adoptado un enfoque bayesiano para entender lo plausibles que son diversas hipótesis con esos datos, en lugar de lo probables que son los datos con la hipótesis nula (es decir, si no hay ningún efecto). Las principales herramientas que usamos en R son blavaan⁴ y rstanarm.⁵ Podemos probar la probabilidad que tiene un efecto de ser sustancial o drástico, y no simplemente menor. Para evaluar un modelo nos decantamos por la parsimonia. Esto supone comenzar con un modelo muy simplista e ir añadiendo complejidad hasta que deje de estar justificada. Por ejemplo, predecimos que el rendimiento empresarial es el resultado de la interacción entre el rendimiento del envío del software y el rendimiento operativo.

Nuestro modelo simplista no incluye la interacción:

Rendimiento empresarial ~ Rendimiento del envío de software + Rendimiento operativo

Nuestro segundo modelo añade la interacción:

Rendimiento empresarial ~ Rendimiento del envío de software + Rendimiento operativo + Rendimiento del envío de software x Rendimiento operativo

A partir de las recomendaciones que se hacen en "Regression and other stories"⁶ y "Statistical Rethinking",⁷ utilizamos la validación cruzada dejando uno fuera (LOOCV) y el criterio de información de aplicación generalizada de Watanabe-Akaike⁸ para determinar si la complejidad adicional es necesaria.

⁴ Merkle, Edgar C. y Yves Rosseel. "blavaan: Bayesian structural equation models via parameter expansion", borrador de artículo arXiv, 2015. <https://doi.org/10.48550/arXiv.1511.05604>

⁵ Goodrich, Ben, Jonah Gabry, Imdad Ali y Sam Brilleman. "rstanarm: Bayesian applied regression modeling via Stan". Versión 2 del paquete de R, n.º 1 (2020).

⁶ Gelman, Andrew, Jennifer Hill y Aki Vehtari. Regression and Other Stories (Cambridge University Press, 2020).

⁷ McElreath, Richard. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC, 2018.

⁸ Aki Vehtari, Andrew Gelman y Jonah Gabry. "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC". Statistics and Computing, 27, 2017. 1413-1432. <https://doi.org/10.1007/s11222-016-9696-4>

Paso 7: informar de las conclusiones

A continuación revisamos estos resultados como equipo. Este año hemos pasado unos días juntos en Boulder, Colorado, sintetizando los datos con la experiencia de expertos en la materia. Lo repetimos en cada capítulo del informe, hipótesis a hipótesis. La interpretación de los datos siempre conlleva riesgos originados por los giros, la especulación, las anécdotas y los saltos. Estos riesgos se mitigaron al contar en la misma sala con personas con formación diversa que fomentaban la puesta en duda, la divergencia, los puntos de vista singulares y la curiosidad.⁹

Los resultados ya estaban listos, así que los autores del informe se retiraron a sus respectivos rincones del mundo y empezaron a escribir. Durante el proceso de redacción, se consultó a editores y expertos en la materia. Estas perspectivas fueron vitales para ayudarnos a comunicar nuestras ideas. La persona responsable de analizar estos datos fue la encargada de asegurarse de que nada de lo que decíamos se desviaba de lo que decían los datos.

Los capítulos se agruparon con un diseño cohesionado que realizaron nuestros brillantes partners de diseño BrightCarbon.¹⁰

Paso 8: sintetizar las conclusiones con la ayuda de la comunidad

Contamos con la participación de la comunidad para saber cómo aprovechar e interpretar estas conclusiones. Tratamos de hacer recomendaciones específicas, pero hay innumerables implementaciones que un equipo puede probar con nuestros resultados. Por ejemplo, la arquitectura con bajo acoplamiento parece ser una práctica beneficiosa según los resultados que medimos. Pero no existe una sola manera de establecer una arquitectura con bajo acoplamiento. La única forma de mejorar constantemente es generar estrategias como comunidad e intercambiarlas. Nuestro mapa del mundo es una interpretación y abstracción del territorio y el contexto en el que tu equipo, tu empresa y tú trabajáis.

Para participar en la comunidad mundial de prácticas de DORA, visita el sitio de DORA Community (<https://dora.community>).

⁹ Stasser, G. y Titus, W. (1985). "Pooling of unshared information in group decision making: Biased information sampling during discussion". Journal of Personality and Social Psychology, 48(6), 1985. 1467-1478. <https://doi.org/10.1037/0022-3514.48.6.1467>

¹⁰ <https://www.brightcarbon.com/>

Datos demográficos y firmográficos

¿Quién ha completado la encuesta?

El programa de investigación de DORA lleva casi una década estudiando las competencias, prácticas y métricas de empresas tecnológicas con un alto rendimiento. En él han participado más de 36.000 profesionales que trabajan en empresas de diferentes tamaños y sectores. Gracias por vuestras valiosas aportaciones. Este año, cerca de 3000 profesionales de diversos sectores de todo el mundo han compartido sus experiencias para ayudarnos a conocer mejor los factores que impulsan a las empresas tecnológicas con un alto rendimiento.

Para las preguntas de carácter demográfico y firmográfico de este año se ha tenido en cuenta la investigación realizada por Stack Overflow. Más de 70.000 encuestados participaron en la encuesta a desarrolladores de Stack Overflow en el 2022.¹ En esa encuesta no se incluyó a todos los profesionales técnicos, por numerosas razones, pero es lo más parecido a un censo del mundo de los desarrolladores. Al hacernos una idea de la población a partir de esa encuesta, podemos localizar el sesgo de respuesta en nuestros datos y conocer cuánto queremos generalizar nuestras conclusiones. Esos datos y las preguntas de

carácter demográfico y firmográfico que se formularon en la encuesta a desarrolladores de Stack Overflow están bien elaborados, por lo que merece la pena tomarlos en consideración. En relación con la encuesta a desarrolladores de Stack Overflow, nuestro conjunto de muestras incluye una mayor proporción de mujeres, personas con discapacidad y participantes que trabajan en empresas más grandes. Nuestro conjunto de muestras es similar al de Stack Overflow en lo que respecta a raza y etnia.

En la encuesta de este año se multiplicó por 3,6 el número de encuestados orgánicos en comparación con el 2022.

En la encuesta de este año se multiplicó por
3,6
el número de
encuestados orgánicos en
comparación con el 2022

¹ <https://survey.stackoverflow.co/2022#overview>

Datos demográficos

Sexo

En comparación con el 2022, la proporción de mujeres encuestadas en la muestra de este año es más pequeña (12 % frente al 18 %).

Sexo	Porcentaje de encuestados
Prefieren no decirlo	3 %
Terminología propia	2 %
Mujer	12 %
Hombre	81 %

Discapacidad

Identificamos la discapacidad en seis dimensiones que siguen las pautas del Washington Group Short Set.² Este es el quinto año que hemos preguntado sobre discapacidad. El porcentaje de personas con discapacidad disminuyó del 11 % en el 2022 al 6 % en el 2023.

Discapacidad	Porcentaje de encuestados
Ninguna de las discapacidades	87 %
Sí	6 %
Prefieren no decirlo/ Sin respuesta	7 %

Infrarrepresentación

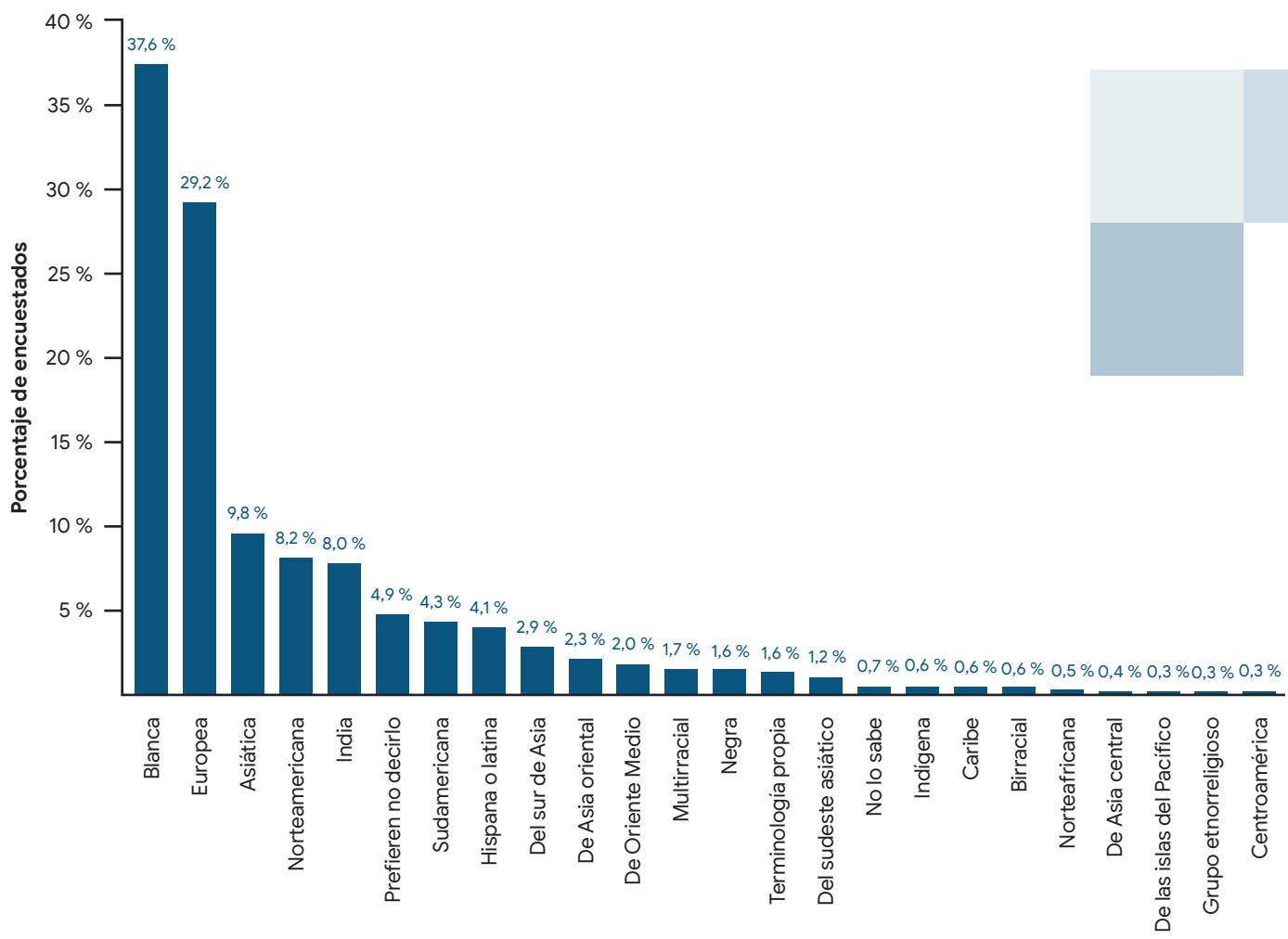
Identificarse como miembro de un grupo infrarrepresentado puede hacer referencia a la raza, sexo u otra característica. Este es el sexto año que hemos preguntado sobre la infrarrepresentación. El porcentaje de personas que se identifican como pertenecientes a un grupo infrarrepresentado ha disminuido ligeramente, del 19 % en el 2022 al 15 % en el 2023.

Infrarrepresentación	Porcentaje de encuestados
No	77 %
Sí	15 %
Prefieren no responder	7 %

² <https://www.washingtongroup-disability.com/question-sets/wg-short-set-on-functioning-wg-ss/>

Raza y etnia

Usamos las preguntas de la encuesta a desarrolladores de Stack Overflow del 2022.³ Como señalamos anteriormente, nuestro conjunto de muestras es similar, aunque tenemos una proporción menor de europeos, lo que supone una variación significativa.



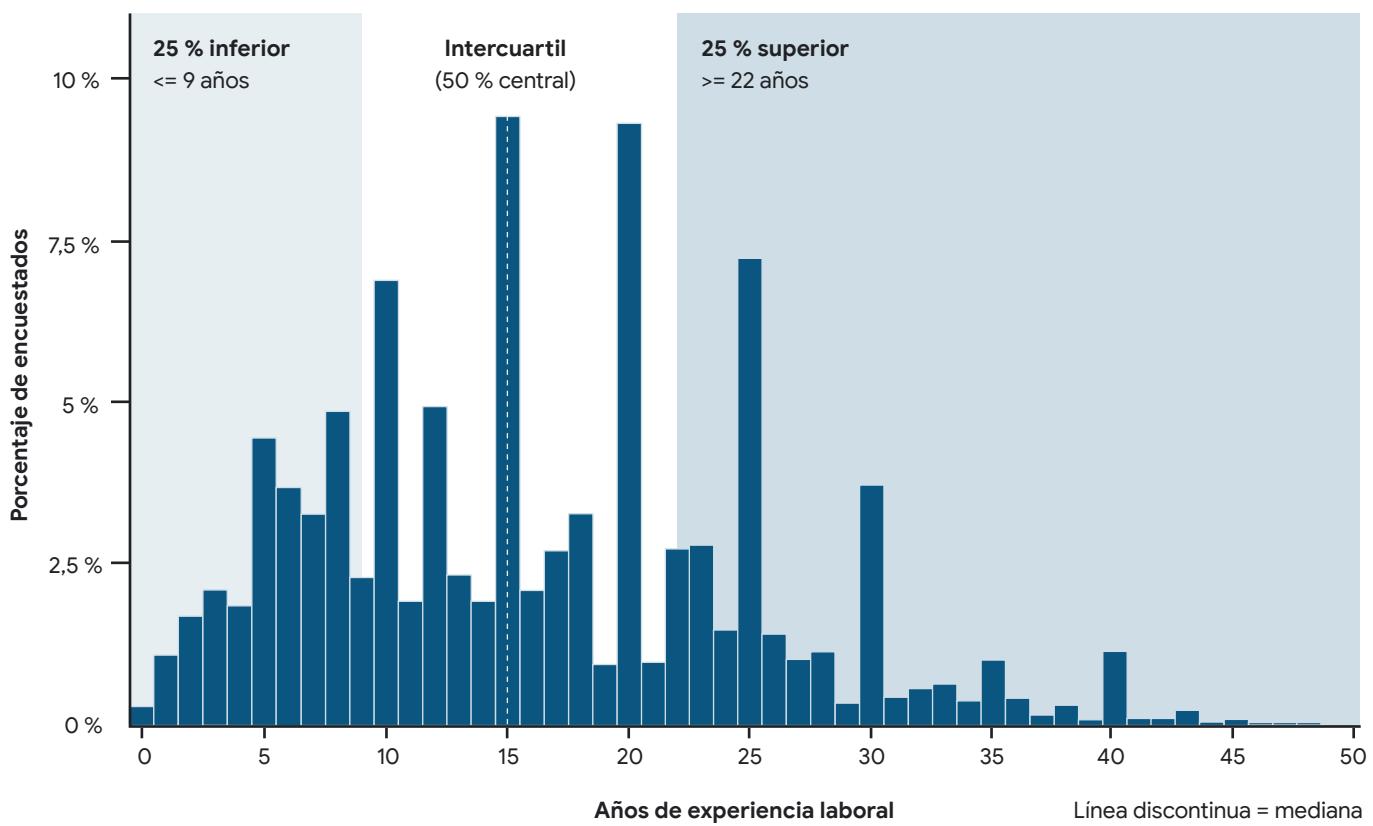
³ <https://survey.stackoverflow.co/2022#overview>

Años de experiencia

Experiencia laboral

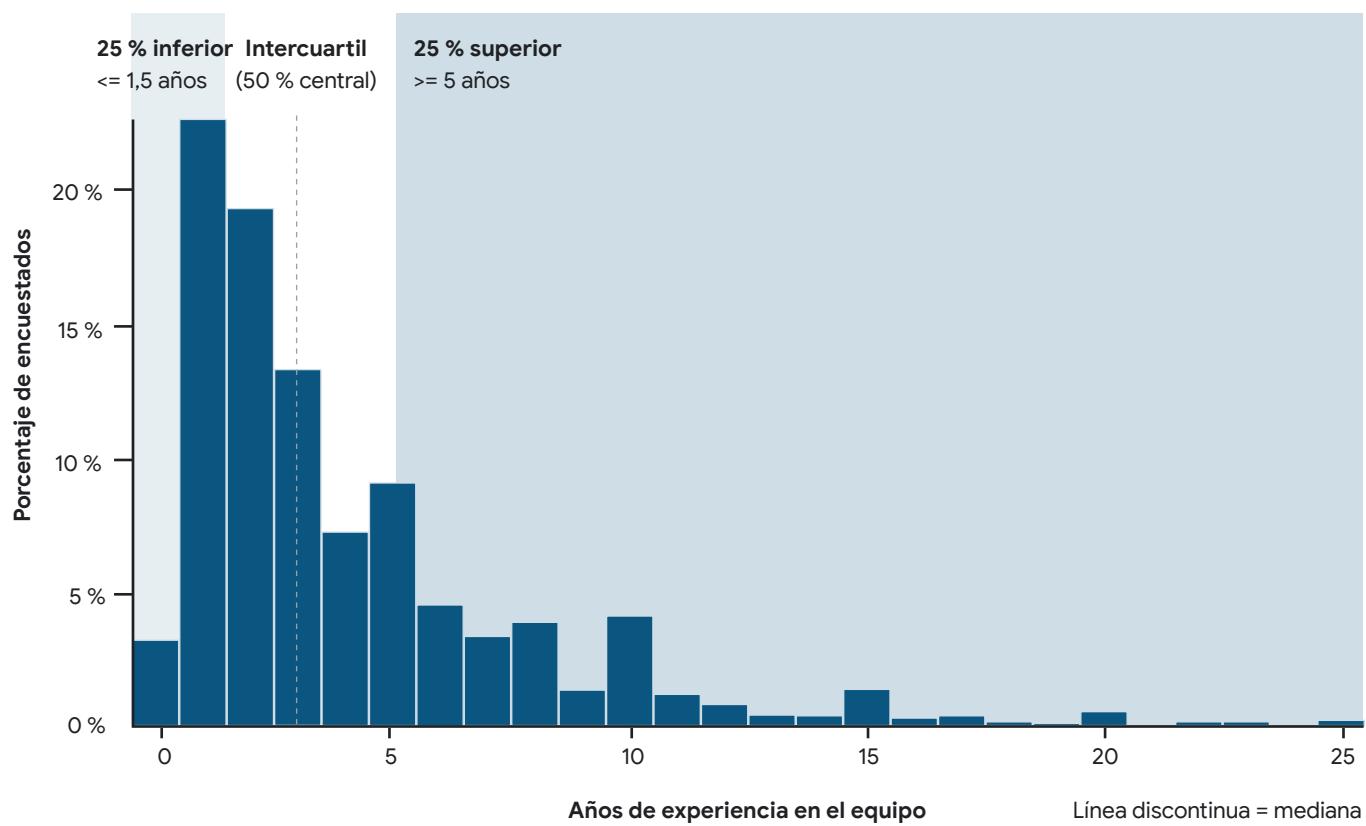
Queríamos saber cuánto tiempo llevaba una persona trabajando, así que preguntamos: "¿Cuántos años de experiencia laboral tienes?". En resumen, estamos recopilando datos de un grupo de profesionales con bastante experiencia. Es decir, el 50 % de nuestros encuestados tenía 15 años o más de experiencia.

El 25 % tenía más de 22 años de experiencia, y el 25 %, menos de 9. A posteriori, nos hemos dado cuenta de no es tan evidente lo que una persona considera "trabajar".



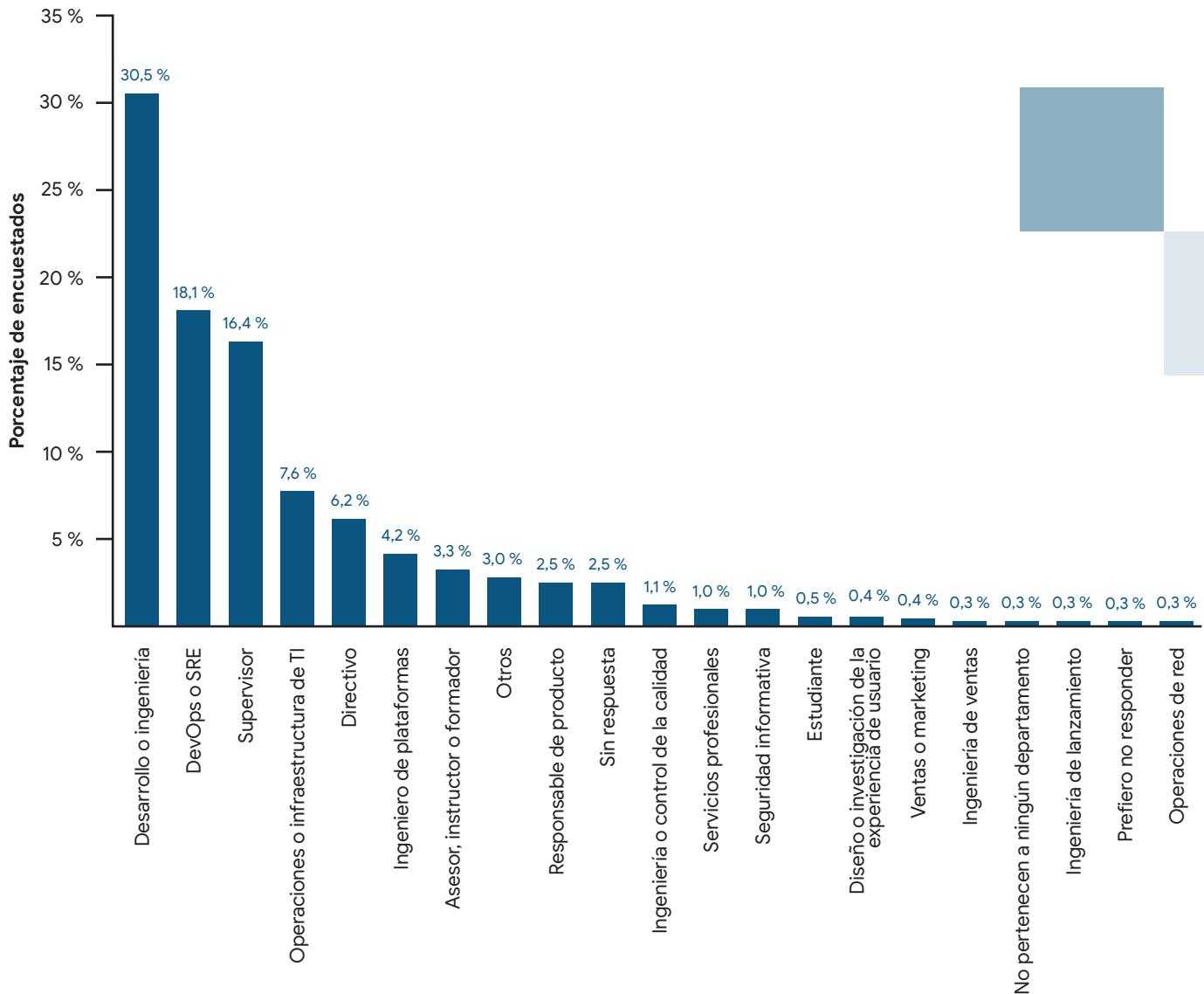
Experiencia en el equipo

A veces trabajar en un equipo nuevo es como empezar de cero, como un pequeño reinicio. Queríamos estudiar esto más en profundidad y para ello preguntamos lo siguiente: "¿Cuántos años llevas trabajando en tu equipo actual?". A pesar de tener una experiencia laboral aparentemente alta, muchos encuestados son nuevos en sus equipos. El 50 % de los encuestados llevaba en su nuevo equipo menos de 3 años, el 25% llevaba un año y medio o menos, y solo el 25 % llevaba cinco años o más. ¿Estos datos reflejan una mentalidad de mejora continua en nuestros encuestados? ¿Reflejan los cambios y la inestabilidad de la economía?



Puesto

El 72 % de los encuestados trabaja en equipos de desarrollo o ingeniería (el 30 %), en equipos de DevOps o SRE (el 18 %), en equipos de operaciones o infraestructura de TI (el 8 %) o son supervisores (el 16 %). En el 2022, esos puestos equivalían al 85 % de los encuestados. El descenso de encuestados en esos cuatro puestos sugiere que hemos sido capaces de llegar a más personas en diferentes puestos. La proporción de equipos de operaciones o infraestructura de TI (el 8 %) regresa a los niveles del 2021 (el 9 %) tras la desviación del 2022 (el 19 %).



Datos firmográficos

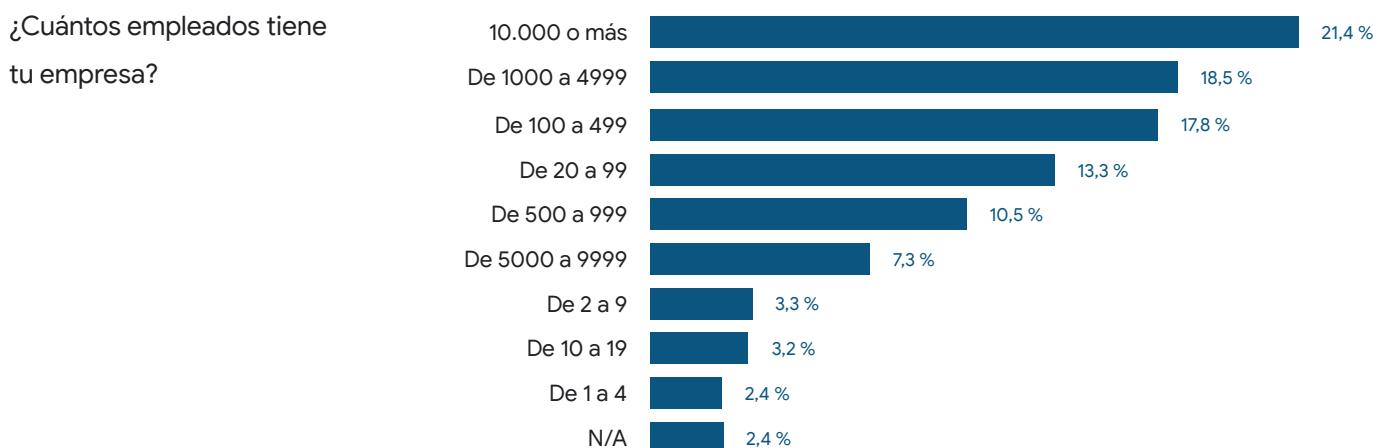
Sector

Porcentaje de encuestados



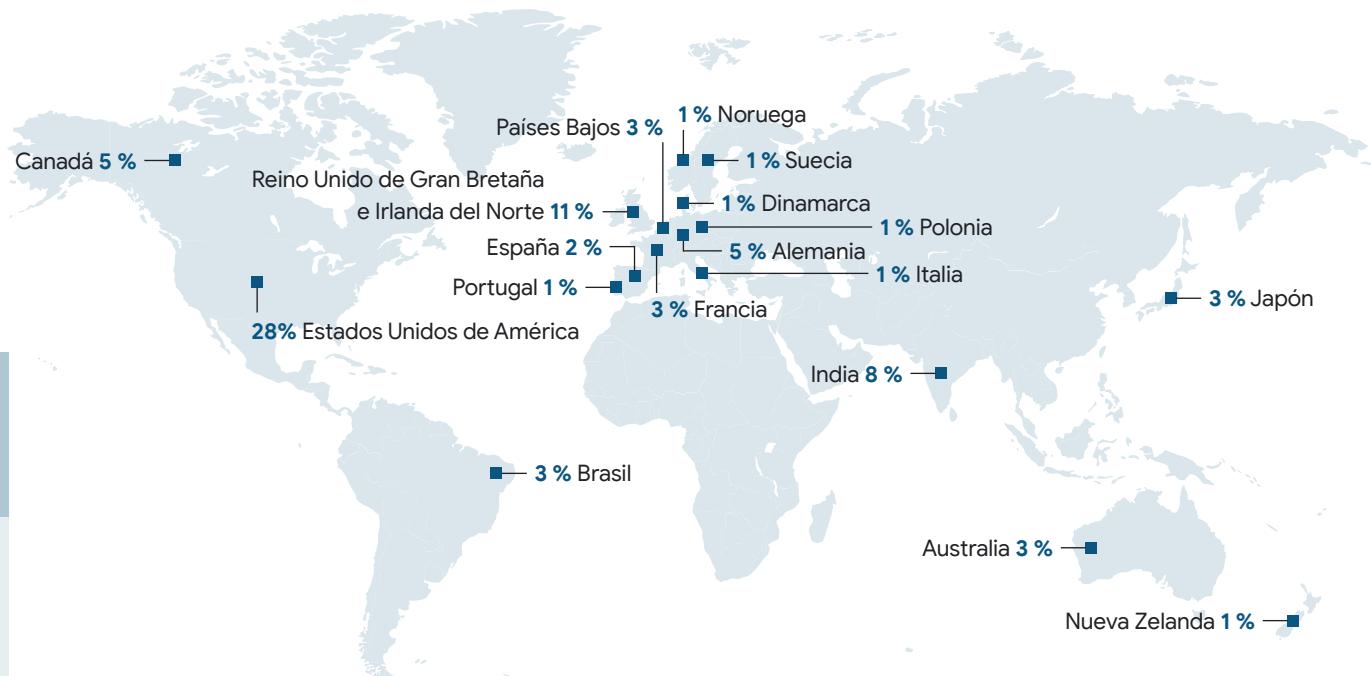
Número de empleados

Porcentaje de encuestados



País

Nos encanta ver que en la encuesta participan personas de todo el mundo. ¡Gracias a todos!



País

Estados Unidos	Dinamarca	Lituania	Túnez	Bangladés	Guatemala
Reino Unido	Suiza	Tailandia	Uruguay	República Dominicana	Honduras
India	Austria	Hungría	Afganistán	Ghana	Letonia
Canadá	Kenia	Israel	Argelia	Región Administrativa Especial de Hong Kong	Líbano
Alemania	Sudáfrica	Vietnam	Egipto	Kazajistán	Luxemburgo
Australia	Argentina	EAU	Estonia	Myanmar/Birmania	Maldivas
Brasil	Chequia	Bulgaria	Islandia	Arabia Saudí	Malta
No aplicable	Bélgica	Croacia	Irán	Somalia	Mauricio
Países Bajos	Colombia	Ecuador	Nigeria	Sudán	Mongolia
Japón	Finlandia	Indonesia	Perú	Uganda	Marruecos
Francia	Irlanda	Filipinas	Eslovaquia	Albania	Nepal
España	China	Armenia	Eslovenia	Bahamas	Catar
Suecia	Rumanía	Georgia	Corea del Sur	Bielorrusia	Macedonia del Norte
Italia	Singapur	Grecia	Sri Lanka	Bolivia	
Nueva Zelanda	México	Malasia	Andorra	Camboya	Trinidad y Tobago
Polonia	Turquía	Pakistán	Angola	Costa Rica	República Unida de Tanzania
Noruega	Ucrania	Federación de Rusia	Antigua y Barbuda	Yibuti	
Portugal	Chile	Serbia	Baréin	El Salvador	Zimbabwe

Modalidad de trabajo

Situación laboral

El 88 % de los encuestados son empleados a jornada completa y el 10 % son contratistas. Algunos contratistas señalan experiencias muy diferentes a las de los empleados a jornada completa.

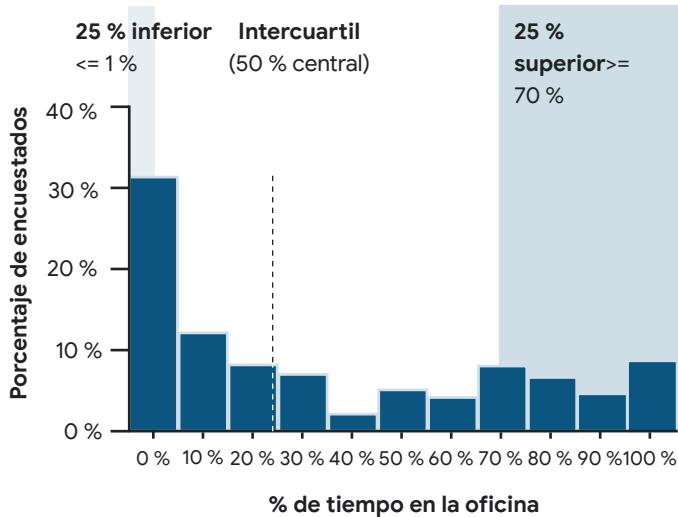
Contrato	Porcentaje de encuestados
Empleado a jornada completa	88 %
Contratista a tiempo completo	8 %
Empleado a tiempo parcial	2 %
Contratista a tiempo parcial	2 %

* Con respecto a la aplicación o servicio principal en el que trabajas, ¿cómo describirías tu situación laboral con la empresa dueña de la aplicación o el servicio?

La diferencia en la experiencia puede deberse a su integración en el equipo. Algunos contratistas declaran estar integrados en el equipo con el que trabajan. Esto significa que trabajan mano a mano con los miembros del equipo a diario y creen que la diferencia con los empleados a jornada completa es mínima. El 70 % de los encuestados que son contratistas están muy de acuerdo o de acuerdo con la afirmación de que están integrados en su equipo.

Ubicación

El patrón de respuesta de este año indica que, a pesar de las presiones por volver a la oficina, trabajar desde casa sigue siendo una realidad para muchos trabajadores. Cerca del 33 % de los encuestados trabaja casi exclusivamente desde casa (menos del 5 % del tiempo en la oficina). El 63 % de los encuestados trabaja desde casa más de lo que lo hace desde la oficina. Para el resto de encuestados, el trabajo híbrido sería la modalidad más común. Eso es lo que sugieren las respuestas, ya que el 75 % de los encuestados pasa menos del 70 % de su tiempo en la oficina. No hay muchas personas que sientan un vínculo fuerte hacia su oficina Solo el 9 % de los encuestados está en la oficina más del 95 % del tiempo.



Los modelos

Introducción

Siempre habíamos creado un único modelo gigante.

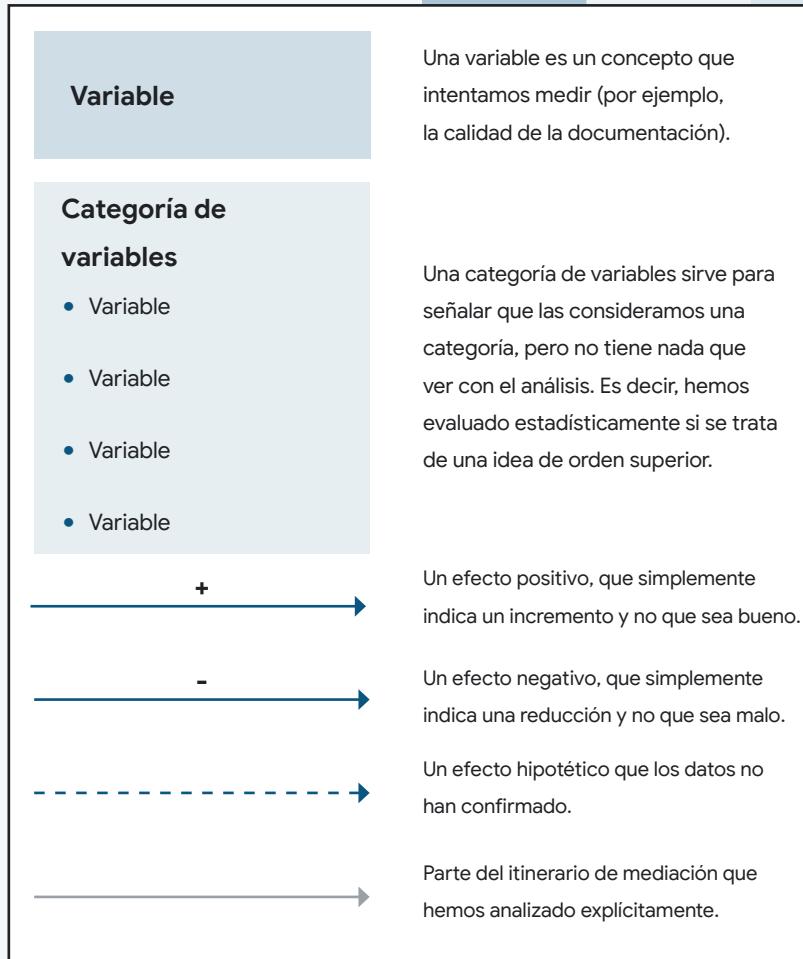
Este año hemos decidido desglosarlo en varios modelos por las siguientes razones:

- Los modelos grandes pueden ser difíciles de manejar con rapidez. Cada variable que se añade modifica el funcionamiento del modelo. Esto puede provocar cálculos imprecisos y complica la tarea de localizar el motivo del cambio.
- Este año hemos creado nuestras hipótesis sección a sección. Por eso tiene sentido que simplemente creemos un modelo para cada sección.
- La ventaja que ofrece un modelo gigante a la hora de calcular el efecto que X tiene sobre Y no resulta evidente. Para entender el efecto de X sobre Y, hemos usado grafos acíclicos dirigidos para que nos ayuden a saber qué covariante debemos incluir o no en el modelo.
- Con la cantidad de hipótesis que hemos estudiado este año, sería muy difícil para el lector encontrar sentido al modelo gigante. Imagina combinar todas las visualizaciones que aparecen a continuación en una sola.



¿Cómo leo estos diagramas?

Una vez que sepas leer los diagramas, verás que son herramientas eficientes para transmitir una gran cantidad de información.



Advertencia: Los modelos son sumas generales.

Para facilitar la lectura hemos categorizado algunas variables juntas. Esta estrategia de categorización hace posible que la flecha que va a una categoría de variables, desde una categoría de variables o ambas opciones, sea el patrón general de resultados, aunque puede que no funcione así para cada variable de la categoría. Por ejemplo, el intercambio de conocimientos influye positivamente en la mayoría de los resultados. Por eso, dibujamos una flecha con el signo más (+) encima desde el intercambio de conocimientos hasta la categoría de variables de los resultados clave. Sin embargo, el intercambio de conocimientos no tiene un efecto positivo en el rendimiento del envío de software. Para ver el resto, consulta los capítulos correspondientes.

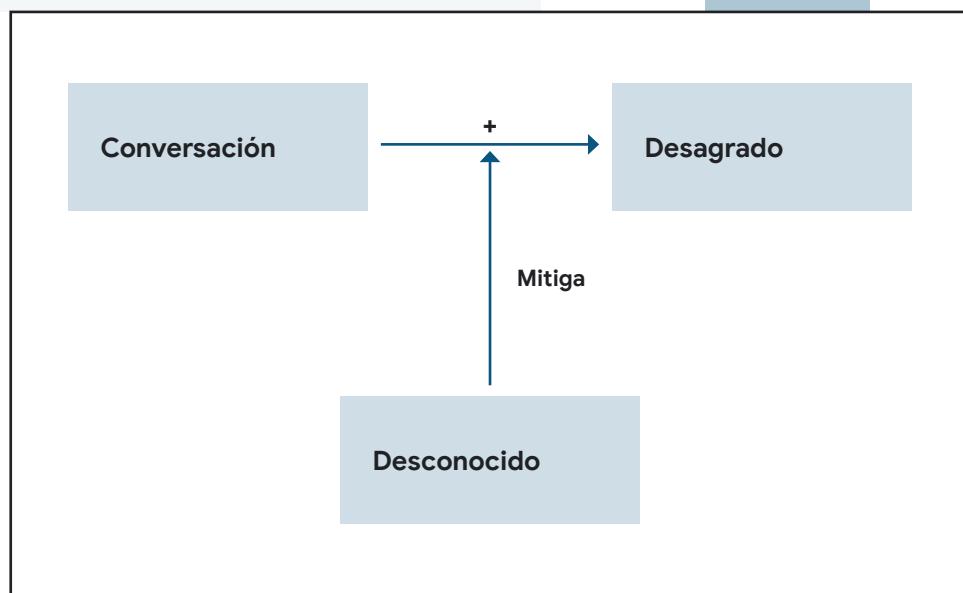
Ejemplo de moderación

La moderación es un concepto difícil de entender en estadística, pero, en el mundo real, la moderación equivale a decir "depende". Veamos un pequeño ejemplo para aclarar el concepto de moderación en el contexto de este informe.

En la tercera temporada de la serie Larry David, este dice: "No me gusta hablar con las personas que conozco, pero con los desconocidos no tengo ningún problema". Esta afirmación nos permite crear un diagrama rápido que podemos comentar:

"No me gusta hablar con las personas que conozco, pero con los desconocidos no tengo ningún problema".

Larry David



Este diagrama muestra que, para Larry, la conversación tiene un efecto positivo en el desagrado. Positivo en este caso simplemente significa un incremento, no que sea necesariamente algo bueno. Esto se indica con una línea negra continua entre conversación y desagrado con la flecha apuntando a desagrado. La flecha sugiere que creemos que el flujo causal va de conversación a desagrado. Por lo que sabemos, las conversaciones tienden a causar desagrado a Larry.

Lo segundo que hay que tener en cuenta es que "desconocido" (aquí se usa para representar el booleano desconocido sí/no) no apunta a ninguna

otra variable, sino a un efecto, una flecha. Eso significa que pensamos que desconocido no modifica una variable, sino un efecto. Por ello, dibujamos la flecha desde desconocido hasta otra flecha, y no hasta otra variable. Así que, si Larry habla o no con un desconocido, influye en el efecto que tiene la conversación en el desagrado. Es decir, el efecto de la conversación en el desagrado depende de si la persona con la que charla Larry es un desconocido o no. Cuando lo es, el efecto de la conversación es algo con lo que Larry "no tiene ningún problema". Se podría decir que los desconocidos mitigan el desagrado que Larry siente al mantener una conversación.

Una cosa puede moderar a otra de varias formas:

- **Amplifica:** hace que los efectos positivos sean más positivos y los negativos más negativos.
- **Atenúa:** debilita el efecto.
- **Mitiga:** hace que los efectos positivos sean menos positivos y los negativos menos negativos.
- **Invierte:** hace que los efectos positivos sean negativos y los negativos positivos.
- **Modifica:** a veces el efecto sencillamente cambia, pero el patrón no se puede resumir bien en una sola palabra. Esto suele ocurrir cuando las causas son variables categóricas. Por ejemplo, hay sectores que pueden tener un comportamiento diferente en circunstancias distintas, como las fluctuaciones económicas.

Ejemplo de mediación

Al igual que la moderación, analizamos a fondo la mediación. La base de la mediación son las preguntas *por qué* o *cómo*. ¿Por qué el sol hace que una planta crezca más? ¿Por qué comer fresas me permite sentir el sabor dulce? ¿Cómo se mueve el coche al pisar el acelerador? ¿Cómo reducen el malestar los analgésicos? Podemos probar la mediación en la estadística.¹ Eso equivale a poder decir que "parece que el efecto de *X* en *Y* se explica total o parcialmente por *M*". Por ejemplo, el efecto del sol en el tamaño de una planta se explica por la fotosíntesis.

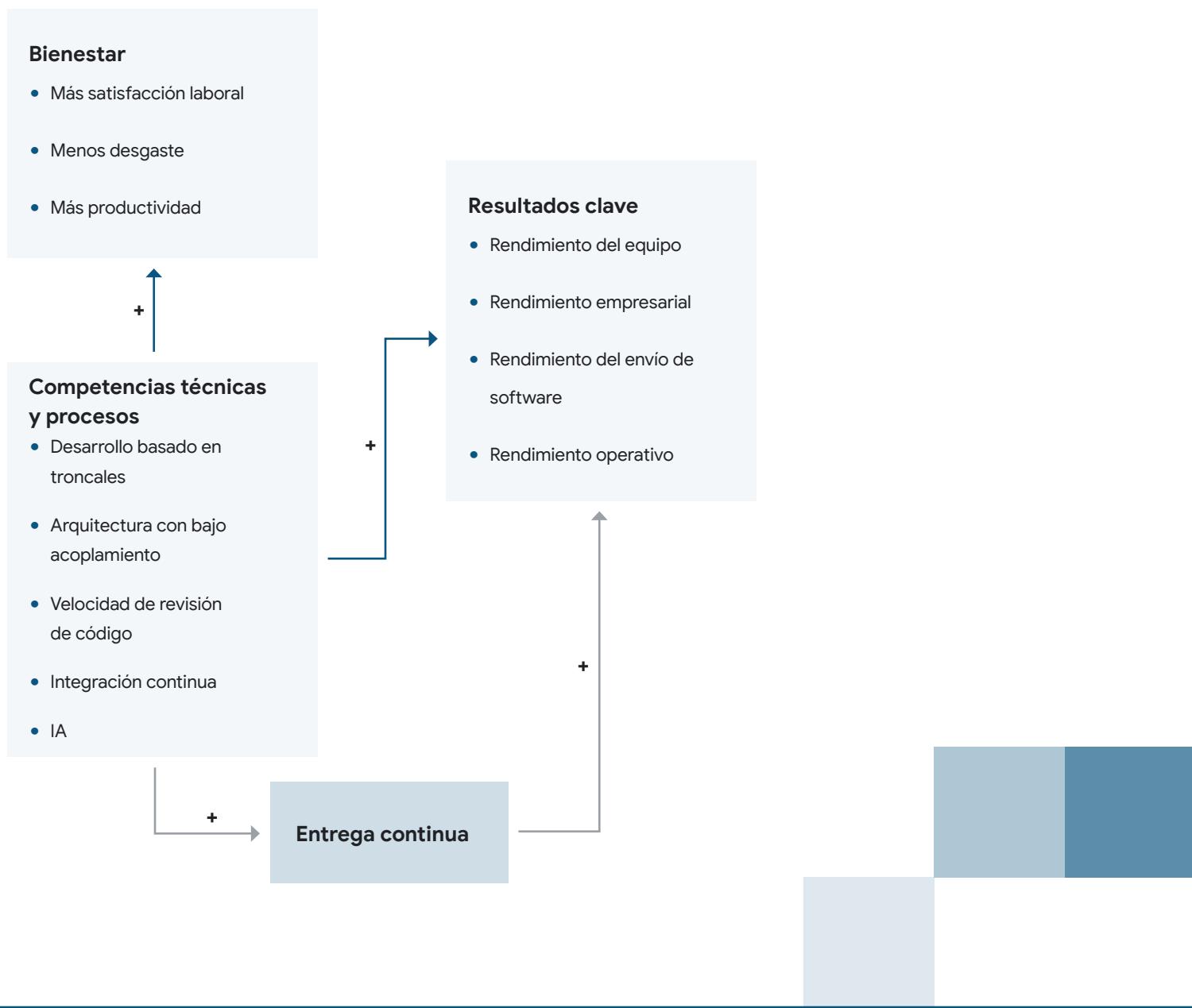


¹<https://lavaan.ugent.be/tutorial/mediation.html>

Modelo del capítulo 3

Las competencias técnicas conducen a un mayor rendimiento

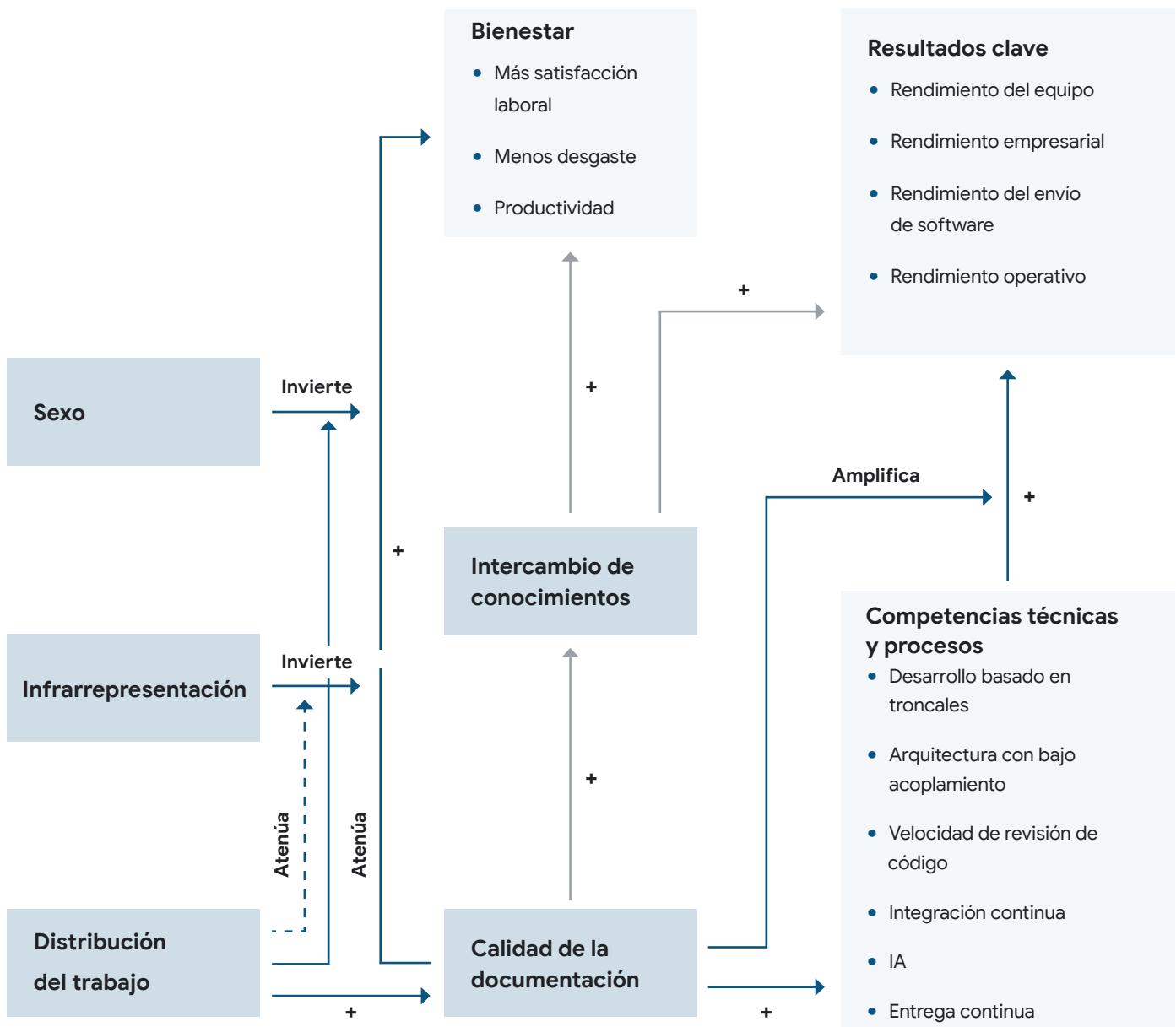
Las competencias técnicas y los procesos tienen un efecto positivo en el bienestar y los resultados clave. El efecto de las competencias técnicas en los resultados clave está mediado en parte por la entrega continua, de tal manera que esas competencias técnicas crean un entorno de entrega continua que tiene su consecuente efecto en los resultados clave.



Modelo del capítulo 4

La documentación es fundamental

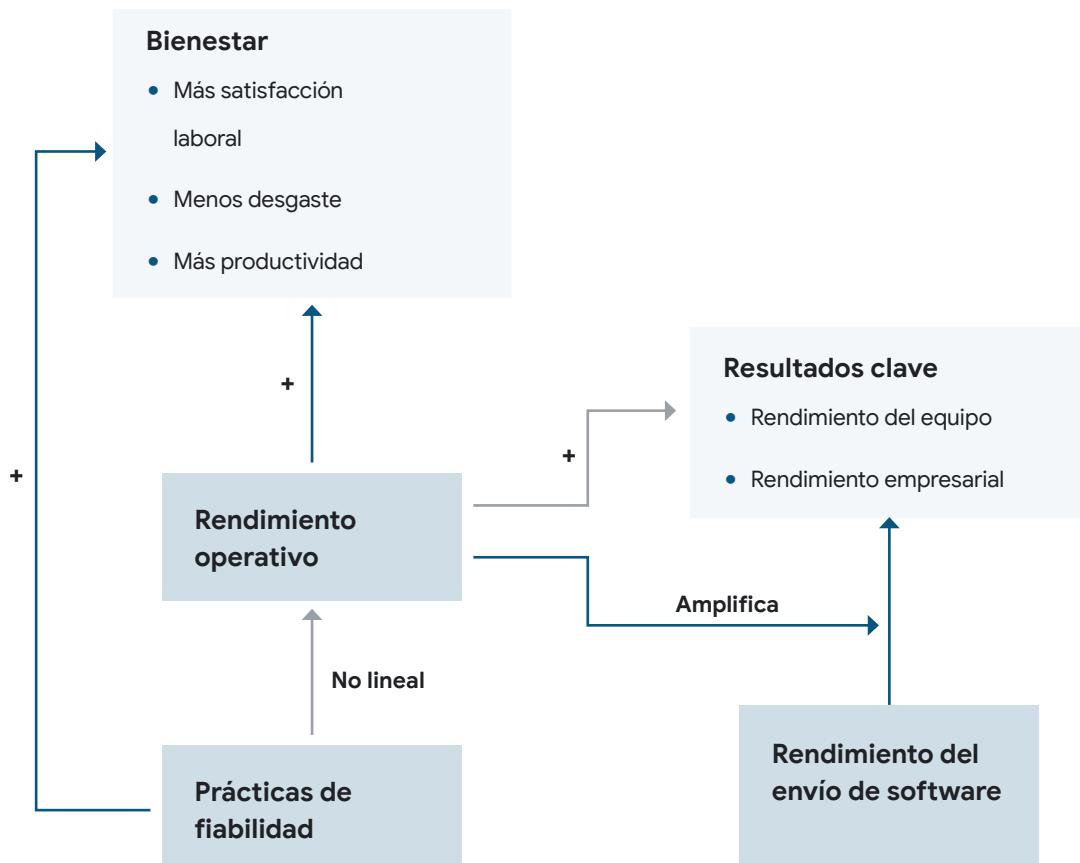
En resumen, la documentación de alta calidad tiene efectos positivos en el bienestar y en los resultados clave. Esto se debe, en parte, a que crea un entorno que posibilita el intercambio de conocimientos. La documentación de alta calidad también ayuda a los equipos a establecer competencias técnicas y procesos. Además, ayuda a que las competencias técnicas y los procesos tengan un efecto todavía mayor en los resultados clave. Por último, la documentación de calidad no produce una mejora del bienestar a todo el mundo. Recomendamos leer la sección correspondiente para consultar los detalles de esta compleja conclusión.



Modelo del capítulo 5

La fiabilidad promueve el rendimiento

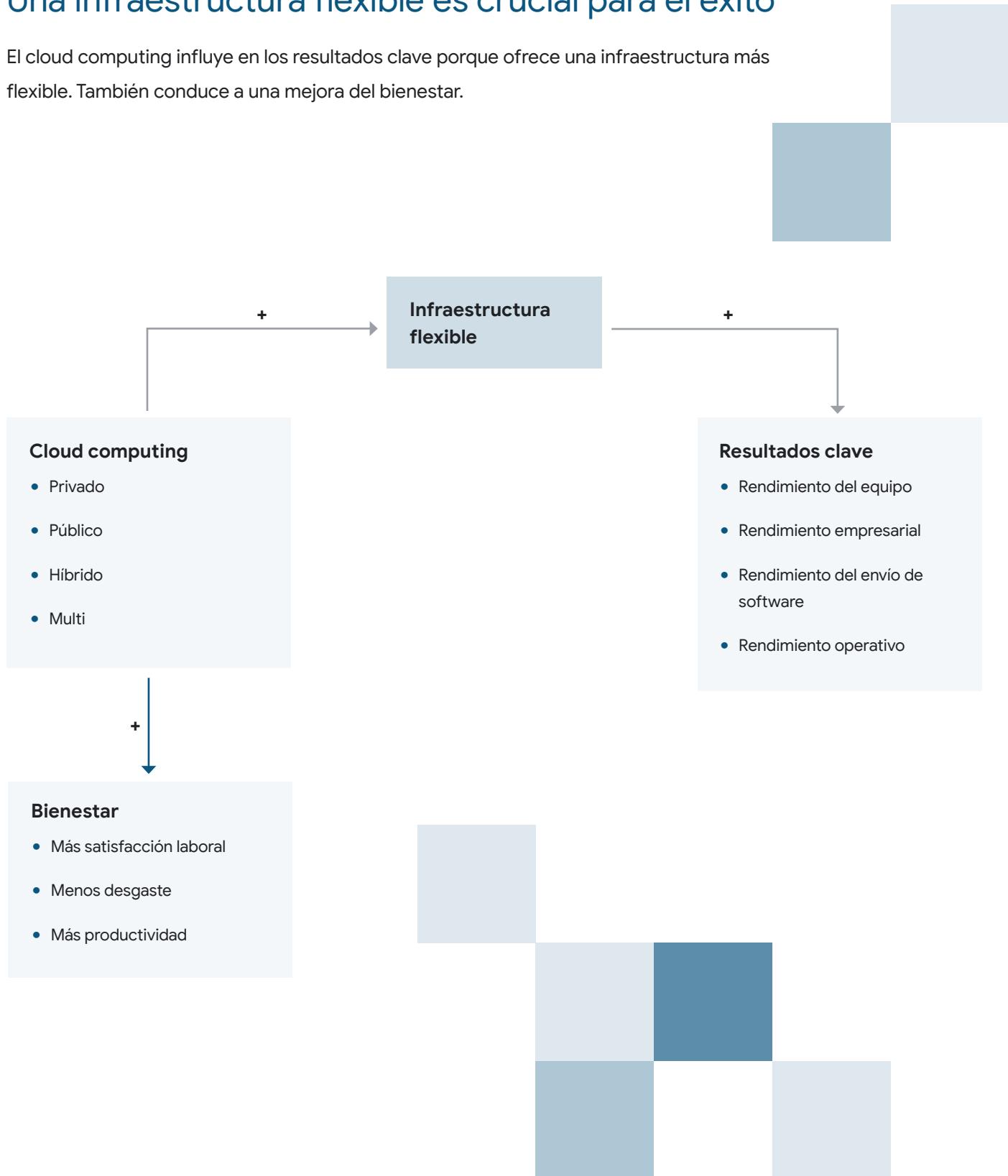
Aquí investigamos el papel fundamental del rendimiento operativo. No solo mejora el bienestar y los resultados clave, sino que también amplifica el efecto del rendimiento del envío de software. Las prácticas de fiabilidad tienen una relación no lineal con el rendimiento operativo. Recomendamos leer este capítulo para entender esos detalles y más.



Modelo del capítulo 6

Una infraestructura flexible es crucial para el éxito

El cloud computing influye en los resultados clave porque ofrece una infraestructura más flexible. También conduce a una mejora del bienestar.

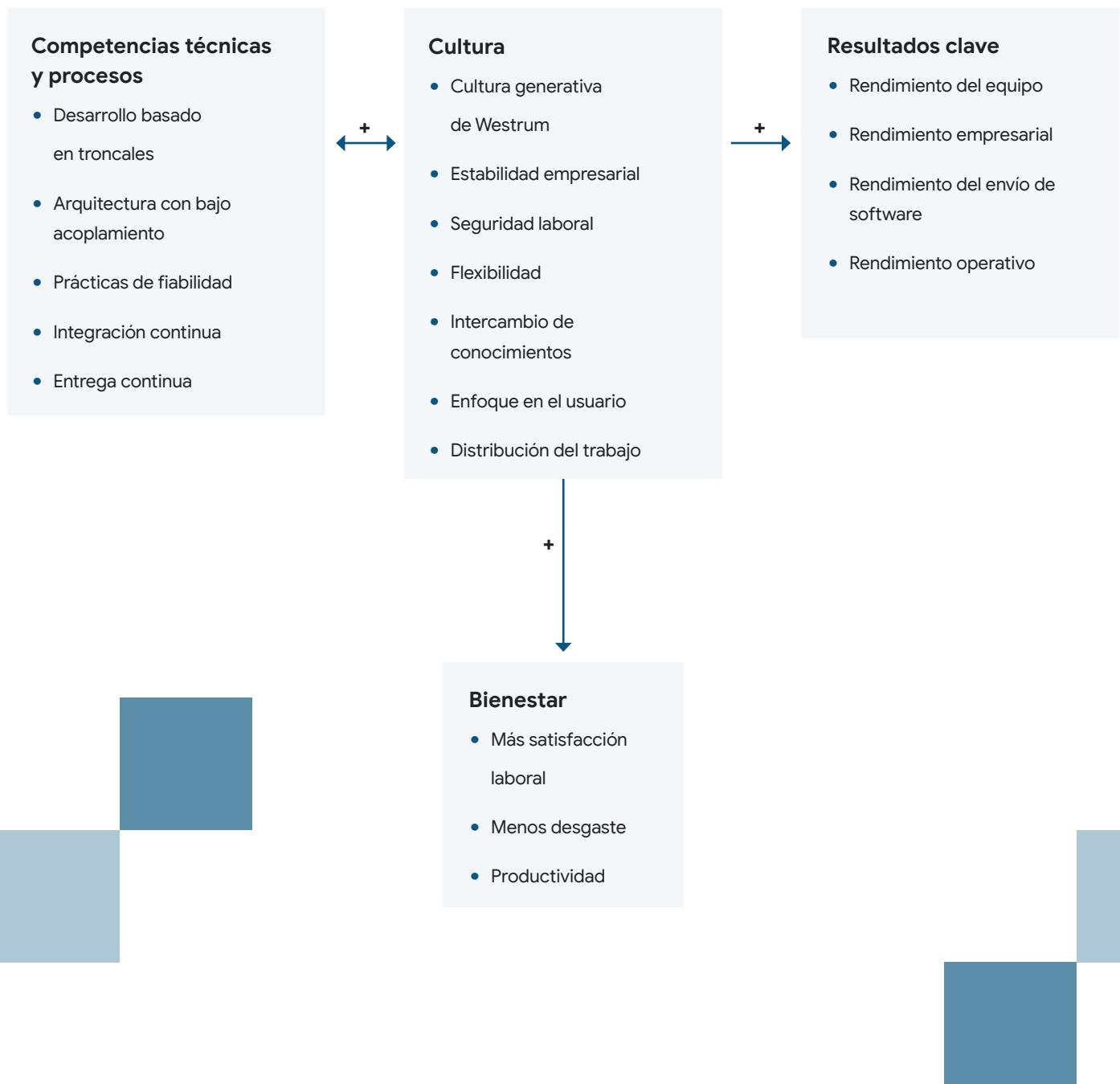


Modelo del capítulo 7

Nada de esto es posible si no se invierte en la cultura

En este diagrama podemos ver que la cultura se encuentra en el centro de muchas cosas.

Observamos que tiene una relación positiva con las competencias técnicas, los resultados clave y el bienestar.

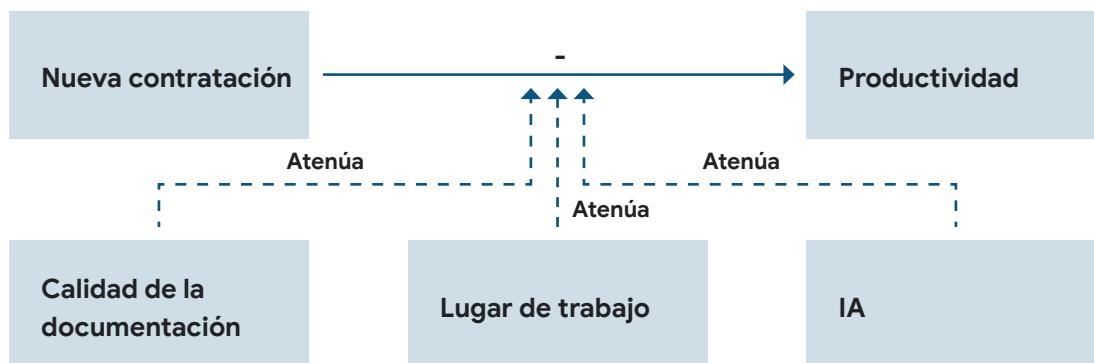
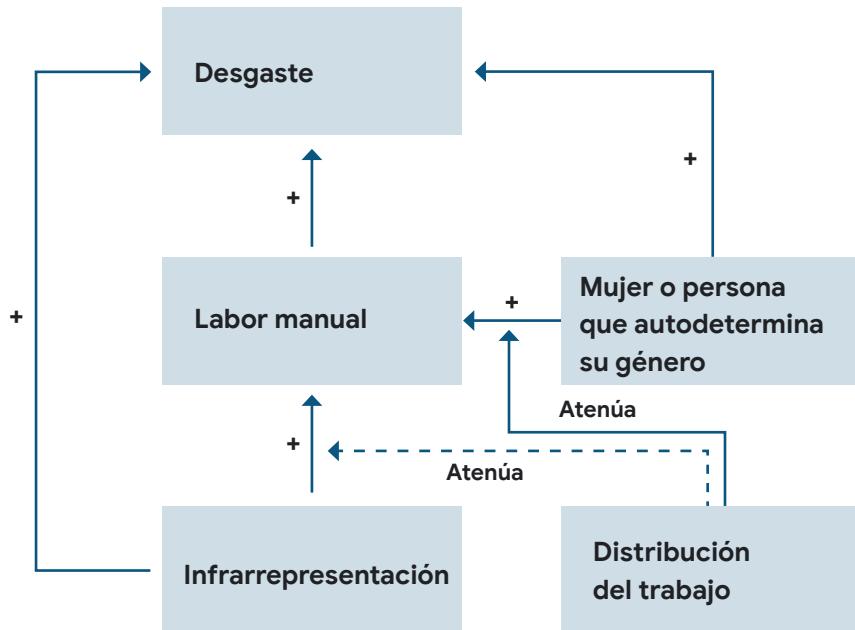


Modelos del capítulo 8

Cómo, cuándo y por qué importa ser quien eres

En esta sección hay dos modelos. Uno investiga por qué y cuándo experimentan niveles más altos de desgaste las personas que se identifican como pertenecientes a un grupo infrarrepresentado y las personas que no se identifican como hombres.

El otro modelo estudia si la documentación de calidad, el lugar de trabajo o la IA pueden mejorar la productividad de las nuevas contrataciones.



Más información

Únete a la DORA Community para debatir, aprender y colaborar a fin de mejorar el rendimiento del envío de software y de las operaciones [DORA.community](https://dora.community)

Haz la encuesta de DevOps de DORA
<https://dora.dev/quickcheck>

Descubre las competencias técnicas, culturales y en procesos que mejoran tanto el rendimiento del envío de software como el empresarial. <https://dora.dev/devops-capabilities/>

Encuentra recursos sobre SRE.
<https://sre.google>
<https://goo.gle/enterprise-roadmap-sre>

Lee el libro *Accelerate: The science behind devops: Building and scaling high performing technology organizations.* IT Revolution.
<https://itrevolution.com/product/accelerate/>

Descubre la constelación adecuada de métricas para tu equipo con el SPACE Framework. "The SPACE of Developer Productivity: There's more to it than you think". <https://queue.acm.org/detail.cfm?id=3454124>

Se han hecho varios estudios de investigación sobre revisiones de código modernas. Aquí tienes algunos:

- "Expectations, Outcomes, and Challenges of Modern Code Review"
<https://dl.acm.org/doi/10.5555/2486788.2486882>

- "Code Reviews - From bottlenecks to Superpowers".
<https://learning.acm.org/techtalks/codereviews>
- "Modern Code Review- A Case Study at Google"
<https://dl.acm.org/doi/10.1145/3183519.3183525>
- "The Choice of Code Review Process: A Survey on the State of the Practice" https://link.springer.com/chapter/10.1007/978-3-319-69926-4_9
- "Investigating the effectiveness of peer code review in distributed software development based on objective and subjective data"
<https://jserd.springeropen.com/articles/10.1186/s40411-018-0058-0>

Lee el libro *The No Club: Putting a Stop to Women's Dead-End Work.* Simon & Schuster.
<https://www.simonandschuster.com/books/The-No-Club/Linda-Babcock/9781982152338>

Publicaciones del programa de investigación de DORA, entre ellas, informes Accelerate State of DevOps anteriores. <https://dora.dev/publications/>

Preguntas frecuentes sobre la investigación y los informes. <http://dora.dev/faq>

Fe de erratas: lee y envía cambios, correcciones y aclaraciones del informe en <https://dora.dev/publications/errata>

Apéndice

Redefinir cómo medimos el rendimiento del envío de software

Este año hemos cambiado la manera de evaluar los fallos por cambios y la recuperación de los fallos.

Hemos modificado la forma en que los encuestados registran su **índice de fallos por cambios**. Los años anteriores ofrecíamos seis opciones (0-15 %, 16-30 %, etc.). Este año hemos facilitado un control deslizante para que puedan seleccionar cualquier valor entre 0 % y 100 %.

Lo hemos cambiado por dos razones:

- El índice de fallos por cambios siempre se ha comportado de una manera un poco distinta a la de las otras tres medidas del rendimiento del envío de software.¹ Tenemos la teoría de que esto podría deberse, en parte, al tamaño de los segmentos. Si la respuesta fuera más precisa, los resultados estadísticos serían mejores. Hemos podido validar esta teoría.
- Tenemos la hipótesis de que los equipos entienden mejor su índice de fallos por cambios ahora que cuando la investigación comenzó hace casi una década. Pero lo cierto es que no tenemos forma de validar esta hipótesis.

Este año hemos preguntado sobre la recuperación de los fallos de dos maneras diferentes:

- Antes hacíamos esta pregunta: "Respecto a la aplicación o servicio principal en los que trabajas, ¿cuánto se suele tardar en restaurar el servicio cuando se produce un incidente o defecto que afecta a los usuarios (por ejemplo, una interrupción o alteración imprevistas del servicio)?".
- Este año hemos añadido los siguientes calificadores a la pregunta (aquí señalamos las diferencias en negrita, pero en la encuesta el formato es uniforme): "Respecto a la aplicación o servicio principal en los que trabajas, ¿cuánto se suele tardar en restaurar el servicio **después de que un cambio en la fase de producción o lanzamiento a los usuarios perjudique el servicio (por ejemplo, lo altere o interrumpa) y, en consecuencia, requiere una solución inmediata (como un paquete de revisiones, una restauración, una corrección inmediata o un parche)**?"

¹ Forsgren, N., Humble, J. y Kim, G. *Accelerate: The Science of Lean Software and DevOps: Building and Scaling High Performing Technology Organizations* (IT Revolution Press, 2018), 37–38.

La forma anterior de preguntar sobre los tiempos de recuperación no daba pie a distinguir entre un fallo iniciado por cambio de software y un fallo iniciado por algo como un terremoto que interrumpe el servicio en un centro de datos. Teníamos la hipótesis de que un lenguaje más preciso nos permitiría comparar tipos de fallos similares y estaría más ajustado estadísticamente a las otras tres medidas de rendimiento del envío de software.

Ahora usamos el término "Tiempo de recuperación tras un despliegue fallido" para distinguir nuestro indicador del más genérico "Tiempo para restaurar" que hemos utilizado en el pasado y que a veces se abreviaba como "MTTR". MTTR causaba cierta confusión en la comunidad: ¿la M hace referencia a media o a mediana? Además, los profesionales que quieren aprender más de los fallos, como los que trabajan en ingeniería de resiliencia, están dejando atrás el MTTR como medida fiable para orientar el aprendizaje y la mejora.²

La nueva pregunta que hemos añadido y la nueva métrica (Tiempo de recuperación tras un despliegue fallido) se ajustan más a la intención de medir el rendimiento del envío de software.

² "Moving Past Simple Incident Metrics: Courtney Nash on the VOID"
<https://www.infoq.com/articles/incident-metrics-void/>



Las matemáticas que respaldan las comparaciones

En este informe se incluyen estadísticas que indican que una medición más alta de una variable concreta implica unas puntuaciones N veces más altas respecto de otro elemento. ¿Qué significa esto? ¿A qué hace referencia? A continuación encontrarás nuestra receta para crear estas fórmulas:

1. Creamos una fórmula a partir del modelo evaluado con técnicas de regresión que tienen en cuenta las posibles rutas de sesgo:

$$\text{Felicidad} \sim 5,64 + 0,19 * \text{luz solar} + 0,14 * \text{temperatura}$$

2. Para la variable que nos interesa, por ejemplo "luz solar", buscamos valores altos y bajos para comparar:

$$\text{media} = 6,3; \text{desviación estándar} = 1,4$$

$$\text{Nivel alto de luz solar} = 1 \text{ desviación estándar por encima de la media} = \text{media} + \text{desviación estándar} = 6,3 + 1,4 = 7,7$$

$$\text{Nivel bajo de luz solar} = 1 \text{ desviación estándar por debajo de la media} = \text{media} - \text{desviación estándar} = 6,3 - 1,4 = 4,9$$

3. Calculamos la media de las covariables. En este ejemplo sería "temperatura", que equivale a 0,24 (temperatura estandarizada).
4. Rellenamos la fórmula del paso 1 con datos de luz solar alta y luz solar baja. Observarás que en la fórmula solo hay un número diferente. Así es cómo conservamos todo lo demás igual y solo aislamos la única diferencia de interés.

Predicción de felicidad con nivel alto de luz solar = $5,64 + 0,19 * 7,7 + 0,14 * 0,24 = 7,1$

Predicción de felicidad con nivel bajo de luz solar = $5,64 + 0,19 * 4,9 + 0,14 * 0,24 = 6,6$

5. Calculamos el ratio:

$$\frac{\text{Predicción de felicidad con nivel alto de luz solar}}{\text{Predicción de felicidad con nivel bajo de luz solar}} = \frac{7,1}{6,6} = 1,1 \text{ vez}$$

6. Este ratio sugiere que unos niveles altos de luz solar implican unos niveles un 10 % más altos de felicidad en comparación con unos niveles bajos de luz solar.

¿Qué es una simulación?

No es que nos inventemos los datos. Utilizamos estadísticas bayesianas para calcular una **distribución posterior** que trata de captar "la frecuencia esperada con la que aparecerán diferentes valores de parámetros".³ La parte de simulación extrae datos más de 1000 veces de esta distribución para observar los valores más creíbles de un parámetro (la media, el peso beta, sigma, intercepción, etc.) a partir de nuestros datos. "Imagina que la distribución posterior es un cubo lleno de valores de parámetros, números como 0,1, 0,7, 0,5, 1. Dentro del cubo, cada valor existe en proporción a la probabilidad de su distribución posterior, de tal manera que los valores cercanos al pico son mucho más habituales que los que están en la cola".⁴

Todo esto equivale a usar simulaciones para estudiar posibles interpretaciones de los datos y hacernos una

idea de cuánta incertidumbre hay. Piensa que cada simulación es como una pequeña IA que no conoce nada más que nuestros datos y algunas reglas, e intenta llenar los huecos en blanco (parámetros) con una suposición fundamentada. Haz esto 4000 veces y obtendrás las suposiciones de 4000 pequeñas IAs para un parámetro determinado. Se puede aprender mucho de estas suposiciones. Puedes conocer la suposición media, entre qué valores se encuentra el 89 % de estas suposiciones, cuántas suposiciones hay por encima de un determinado nivel, cuánta variación se da en las suposiciones, etc. Incluso puedes divertirte combinando suposiciones (simulaciones) entre numerosos modelos.

Al incluir un gráfico con un montón de líneas o una distribución de posibles valores, estamos intentando enseñarte lo que resulta más plausible según nuestros datos y el nivel de incertidumbre que hay.

³ McElreath, Richard. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC, 2018, p. 50

⁴ McElreath, Richard. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC, 2018, p. 52

