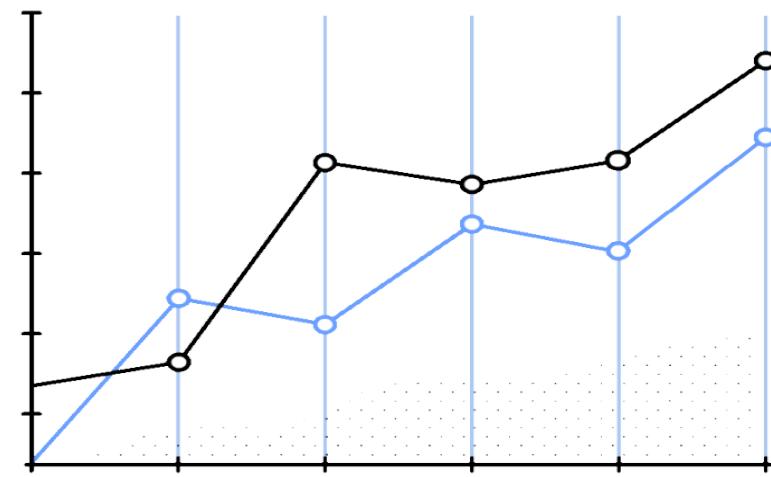


CIENCIA DE DATOS.

¿Sigue siendo el científico de datos el trabajo más atractivo del siglo XXI?

Resumen. Hace diez años, los autores postularon que ser científico de datos era el “trabajo más sexy del siglo XXI”. Una década después, ¿se mantiene la afirmación? El trabajo ha crecido en popularidad y, en general, está bien pagado, y se prevé que el campo experimente un mayor crecimiento que casi cualquier... más.



Hace diez años publicamos el artículo “ Científico de datos: el trabajo más sexy del siglo XXI ”. La mayoría de los lectores casuales probablemente recuerden solo el modificador “más sexy”: un comentario sobre su demanda en el mercado. El puesto era relativamente nuevo en ese momento, pero a medida que más empresas intentaban dar sentido a los grandes datos, **se dieron cuenta de que necesitaban personas que pudieran combinar habilidades de programación, análisis y experimentación**. En ese momento, esa demanda estaba restringida en gran medida al Área de la Bahía de San Francisco y algunas otras ciudades costeras. Las nuevas empresas y las empresas de tecnología en esas áreas parecían querer todos los científicos de datos que pudieran contratar. Sentimos que la necesidad se expandiría a medida que las principales empresas adoptaran tanto el análisis empresarial como nuevas formas y volúmenes de datos.

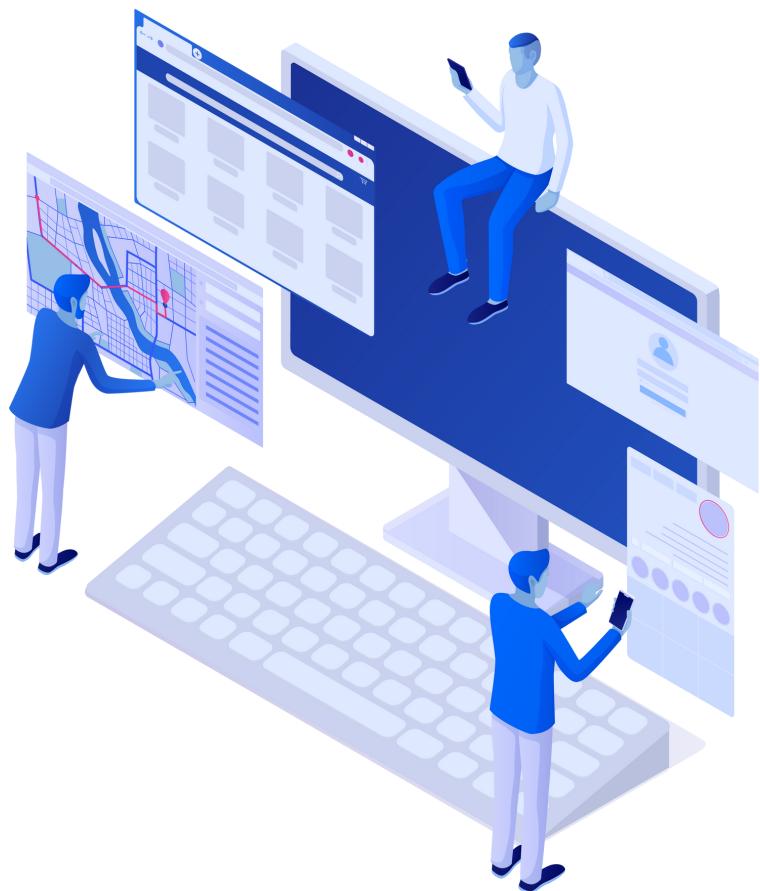
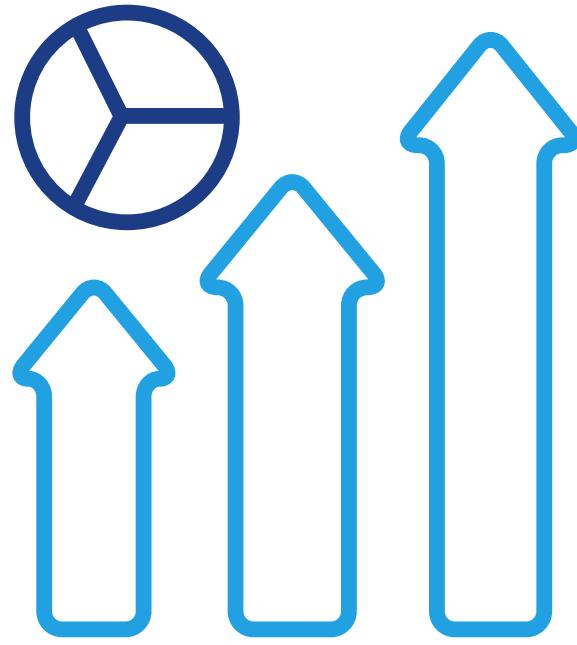
CIENCIA DE DATOS.



En su momento, definimos al científico de datos como “**un profesional de alto rango con la formación y la curiosidad para hacer descubrimientos en el mundo del big data**”. Las empresas comenzaban a analizar datos voluminosos y menos estructurados, como flujos de clics en línea, redes sociales e imágenes y voz. Debido a que aún no había una carrera profesional bien definida para las personas que podían programar y analizar dichos datos, los científicos de datos tenían diversos antecedentes educativos. La calificación más común en nuestra encuesta informal de 35 científicos de datos en ese momento era un doctorado en física experimental, pero también encontramos astrónomos, psicólogos y meteorólogos. **La mayoría tenía doctorados en algún campo científico, eran excepcionales en matemáticas y sabían codificar**. Dada la ausencia de herramientas y procesos en el momento de realizar sus funciones, también eran buenos para la experimentación y la invención.

CIENCIA DE DATOS.

Una década después, el trabajo tiene más demanda que nunca entre empleadores y reclutadores. La IA es cada vez más popular en los negocios, y las empresas de todos los tamaños y ubicaciones sienten que necesitan científicos de datos para desarrollar modelos de IA. Para 2019, las publicaciones para científicos de datos en Indeed habían aumentado un 256 %, y la Oficina de Estadísticas Laborales de EE. UU. predice que la ciencia de datos experimentará un mayor crecimiento que casi cualquier otro campo entre ahora y 2029. El trabajo buscado generalmente se paga bastante bien ; **el salario medio de un científico de datos con experiencia en California se acerca a los 200.000 dólares.**



También quedan muchos de los mismos dolores de cabeza. En nuestra investigación para el artículo original, muchos científicos de datos notaron que pasan gran parte de su tiempo limpiando y discutiendo datos, y eso sigue siendo así a pesar de algunos avances en el uso de la propia IA para mejorar la gestión de datos. Además, muchas organizaciones no tienen culturas basadas en datos y no aprovechan los conocimientos proporcionados por los científicos de

datos. **Ser contratado y bien pagado no significa que los científicos de datos puedan marcar la diferencia para sus empleadores.** Como resultado, muchos se sienten frustrados, lo que lleva a una alta rotación.

CIENCIA DE DATOS.

Aun así, el trabajo ha cambiado, tanto en formas grandes como pequeñas. Se ha institucionalizado mejor, se ha redefinido su alcance, la tecnología en la que se basa ha avanzado a pasos agigantados y **ha crecido la importancia de la experiencia no técnica, como la ética y la gestión del cambio**. Los muchos ejecutivos que reconocen que la ciencia de datos es importante para sus negocios ahora necesitan crear y supervisar diversos equipos de ciencia de datos en lugar de buscar unicornios de científicos de datos. También pueden comenzar a pensar en democratizar la ciencia de datos, sin embargo, aún con la ayuda de científicos de datos mejor institucionalizado.

Mejor institucionalizado

En 2012, la ciencia de datos era una función incipiente incluso en las empresas emergentes orientadas a la IA. **Hoy está bastante bien establecido, al menos en empresas con un gran compromiso con los datos y la IA**. Los bancos, las compañías de seguros, los minoristas e incluso los proveedores de atención médica, e incluso las agencias gubernamentales, tienen importantes grupos de ciencia de datos; las grandes empresas de servicios financieros pueden tener cientos de científicos de datos. La ciencia de datos también ha sido eficaz para abordar crisis sociales, contar y predecir casos y muertes de covid-19, ayudar a abordar desastres meteorológicos e incluso combatir la desinformación y los ataques ciberneticos relacionados con la invasión de Ucrania.



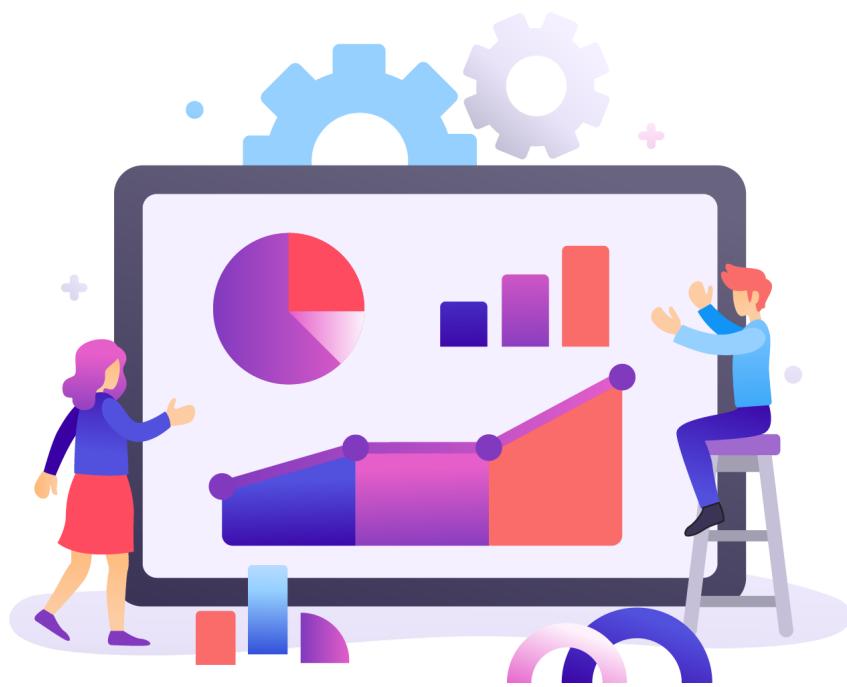
CIENCIA DE DATOS.



Un factor importante que ha facilitado la institucionalización ha sido el **aumento de las ofertas educativas orientadas a la ciencia de datos**. En 2012, efectivamente no había programas de grado en ciencia de datos; Los científicos de datos fueron reclutados de otros campos orientados cuantitativamente. **Ahora hay cientos de programas de grado en ciencia de datos o campos relacionados de análisis e IA**. La mayoría son programas de maestría , pero también hay carreras de pregrado y programas de doctorado en ciencia de datos. También hay una gran cantidad de certificados, ofertas de cursos en línea y campamentos de entrenamiento en campos relacionados con la ciencia de datos. Incluso hay secundaria, cursos y currículos de ciencia de datos. Está claro que **cualquier persona que desee capacitarse en capacidades de ciencia de datos tendrá muchas opciones para hacerlo**. Sin embargo, es poco probable que un solo programa pueda inculcar todas las habilidades necesarias para concebir, construir e implementar análisis, experimentos y modelos de ciencia de datos efectivos y éticos. De hecho, dar sentido a las diversas opciones educativas incluso en una sola institución es un desafío para los futuros científicos de datos y para las empresas que desean emplearlos.

CIENCIA DE DATOS.

Científicos de datos en relación con otros roles



El rol de ciencia de datos ahora también se complementa con una variedad de otros trabajos. La suposición en 2012 era que los científicos de datos podían realizar todas las tareas requeridas en una aplicación de ciencia de datos, desde conceptualizar el caso de uso hasta interactuar con las partes interesadas de negocios y tecnología, desarrollar el algoritmo e implementarlo en producción. Ahora, sin embargo, ha habido una proliferación de trabajos relacionados para manejar muchas de esas tareas, incluido el ingeniero de aprendizaje automático, el ingeniero de datos, el especialista en inteligencia artificial, los traductores de análisis e inteligencia artificial y los gerentes de productos orientados a datos. LinkedIn informó que algunos de estos trabajos son más populares que los científicos de datos en sus informes "Empleos en aumento" para 2021 y 2022 para los EE. UU.



CIENCIA DE DATOS.

Parte de la proliferación se debe al hecho de que ningún titular de un puesto puede poseer todas las habilidades necesarias para implementar con éxito un sistema complejo de inteligencia artificial o análisis. **Cada vez se reconoce más que muchos algoritmos nunca se implementan**, lo que ha llevado a muchas organizaciones a intentar mejorar las tasas de implementación. Además, los desafíos de administrar más sistemas y tecnologías de datos han resultado en un entorno técnico más complejo. Ha habido algunos intentos de certificación de científicos de datos y trabajos relacionados, pero aún no se buscan ni se reconocen ampliamente. **Algunas empresas, como TD Bank , han desarrollado estructuras de clasificación para muchas carreras y habilidades relacionadas con la ciencia de datos**, pero estas no son lo suficientemente comunes en las organizaciones.



Como resultado de esta proliferación de habilidades, las empresas deben identificar todos los diferentes roles necesarios para implementar de manera efectiva los modelos de ciencia de datos en sus negocios y asegurarse de que estén presentes y colaborando en los equipos.

CIENCIA DE DATOS.

Cambios en la tecnología



Una de las razones por las que el trabajo del científico de datos sigue cambiando es porque las tecnologías que usan los científicos de datos están cambiando. Algunas tendencias tecnológicas son la continuación de las direcciones presentes en 2012, como el uso de herramientas de código abierto y el paso al procesamiento y almacenamiento de datos basados en la nube. Pero algunos afectan el núcleo del trabajo de ciencia de datos. Por ejemplo, **algunos aspectos de la ciencia de datos están cada vez más automatizados (usando el aprendizaje automático automático o AutoML)**, lo que puede mejorar la productividad de los profesionales de la ciencia de datos y abrir la posibilidad de "científicos de datos ciudadanos" con solo un poco de capacitación cuantitativa. Estas herramientas automatizadas aún no han disminuido el atractivo de los científicos de datos profesionales, pero es posible que lo hagan en el futuro.

Las empresas deberían comenzar a democratizar el análisis avanzado y la IA dentro de sus organizaciones, confiando en los científicos de datos para garantizar que los modelos desarrollados por los ciudadanos sean precisos y que se empleen todos los datos relevantes.



CIENCIA DE DATOS.

Los científicos de datos se han dado cuenta de que sus modelos pueden "derivar" en entornos comerciales turbulentos como la pandemia de Covid-19, por lo que **hay un nuevo énfasis en monitorear su precisión después de la implementación**. Las herramientas de operaciones de aprendizaje automático, o "MLOps", brindan un monitoreo continuo de los modelos; el reentrenamiento automatizado de modelos desviados apenas comienza a emplearse. Algunas herramientas de AutoML y MLOps incluso prueban el sesgo algorítmico.

Estos desarrollos significan que la codificación, que quizás era el requisito de trabajo más común cuando escribimos hace una década, **es algo menos esencial en la ciencia de datos. Ha migrado a otros trabajos o se está automatizando cada vez más**. (Sin embargo, la limpieza de datos es una excepción notable a esta tendencia). El enfoque clave del trabajo continúa cambiando hacia el modelado predictivo y la capacidad de traducir los problemas y requisitos comerciales en modelos. Estas son actividades colaborativas, pero lamentablemente todavía no existen herramientas excelentes para estructurar y respaldar actividades colaborativas de ciencia de datos.



CIENCIA DE DATOS.



La ética de la ciencia de datos

Un cambio importante en la ciencia de datos durante la última década es que **ahora se reconoce ampliamente la necesidad de una dimensión ética en el campo**, aunque el tema rara vez se mencionó en 2012. El punto de inflexión para la ética de la ciencia de datos fue probablemente la elección presidencial de EE. UU. de 2016, en el que los científicos de datos en las redes sociales (Cambridge Analytica y Facebook en particular) intentaron influir en los votantes y polarizaron aún más la política electoral. Desde entonces, **se ha prestado una atención considerable a las cuestiones del sesgo algorítmico, la transparencia y el uso responsable de la analítica y la IA**.

Algunas empresas ya han establecido grupos y procesos de IA responsables. Una función clave de ellos es educar a los científicos de datos sobre los problemas relacionados con la IA ética. Y hay una mayor regulación que se está instituyendo en respuesta a los lapsos éticos.

CIENCIA DE DATOS.



La ética de la ciencia de datos

Un cambio importante en la ciencia de datos durante la última década es que **ahora se reconoce ampliamente la necesidad de una dimensión ética en el campo**, aunque el tema rara vez se mencionó en 2012. El punto de inflexión para la ética de la ciencia de datos fue probablemente la elección presidencial de EE. UU. de 2016, en el que los científicos de datos en las redes sociales (Cambridge Analytica y Facebook en particular) intentaron influir en los votantes y polarizaron aún más la política electoral. Desde entonces, **se ha prestado una atención considerable a las cuestiones del sesgo algorítmico, la transparencia y el uso responsable de la analítica y la IA**.

Algunas empresas ya han establecido grupos y procesos de IA responsables. Una función clave de ellos es educar a los científicos de datos sobre los problemas relacionados con la IA ética. Y hay una mayor regulación que se está instituyendo en respuesta a los lapsos éticos.

CIENCIA DE DATOS.

Hemos visto tanto la continuidad como el cambio en el rol de la ciencia de datos. Ha tenido un éxito notable en muchos sentidos, y algunos de sus desafíos (proliferación de roles relacionados, la necesidad de una perspectiva ética) resultan en parte de la adopción generalizada de la ciencia de datos. Parece poco probable que disminuya la cantidad de datos, análisis e inteligencia artificial en los negocios y la sociedad, **por lo que el trabajo del científico de datos seguirá creciendo en importancia en el panorama empresarial**.

Sin embargo, también seguirá cambiando. Esperamos ver una diferenciación continua de responsabilidades y roles que alguna vez cayeron en la categoría de científico de datos. Las empresas necesitarán procesos detallados de clasificación y certificación de habilidades para estos diversos trabajos, y deben asegurarse de que todos los roles necesarios estén presentes en proyectos de ciencia de datos a gran escala. Los propios científicos de datos profesionales se centrarán en la innovación algorítmica, pero también deberán ser responsables de garantizar que los aficionados no se desborden. Lo que es más importante, **los científicos de datos deben contribuir a la recopilación adecuada de datos, el análisis responsable, los modelos completamente implementados y los resultados comerciales exitosos**.

