

## **Transcripción del vídeo**

### **¿Qué es la ciencia de datos?**

(Shingai Manjengwa): La ciencia de datos es un proceso, no un evento. Es el proceso de utilizar datos para entender distintas cuestiones... para entender el mundo.

(Rafael B. Da Silva): Para mí es cuando tienes un modelo o una hipótesis de un problema e intentas validarlos con los datos.

(Diana Zárate-Díaz): La ciencia de datos es el arte de descubrir la información útil y las tendencias que se ocultan detrás de los datos.

(Luis O. Martins): Es cuando se convierten los datos en una historia. Para generar información útil, se utiliza la narrativa. Con esta información útil, puede tomar decisiones estratégicas para una empresa o una institución.

(Mandeep Kaur): La ciencia de datos es un campo sobre los procesos y sistemas que se utilizan para extraer datos con varios formatos, que pueden ser estructurados o no estructurados.

(Stephen Sherman): La ciencia de datos es el estudio de los datos. De la misma manera que la biología es la ciencia que estudia los seres vivos y la física es la ciencia que estudia las reacciones físicas. Los datos son reales, los datos tienen propiedades reales y debemos estudiarlas si queremos trabajar con ellos.

(Murtaza Haider): La ciencia de datos implica datos y algunos signos. La definición o el nombre surgió en los 80 y los 90 cuando algunos profesores revisaron los programas de estudios de estadística y consideraron que sería mejor llamarlos ciencia de datos. ¿Pero en qué consiste la ciencia de datos? Para mí, la ciencia de datos es el intento de trabajar con datos y encontrar respuestas a preguntas mediante su exploración. En resumen, tiene que ver más con los datos que con la ciencia. Si tenemos datos y curiosidad, y trabajamos con los datos, los manipulamos y los exploramos... el propio ejercicio de analizar los datos e intentar obtener algunas respuestas, eso es la ciencia de datos.

La ciencia de datos es relevante actualmente porque tenemos toneladas de datos disponibles. Solíamos preocuparnos por la falta de datos. Hoy día, tenemos un diluvio de datos. En el pasado, no teníamos algoritmo y ahora los tenemos. En el pasado, el software era muy caro y ahora es gratuito y de código abierto. En el pasado, no podíamos almacenar grandes cantidades de datos y ahora por una fracción del coste tenemos millones de conjuntos de datos de manera asequible. Tenemos herramientas para trabajar con los datos, variabilidad y capacidad de almacenarlos y analizarlos, todo a un precio económico, con gran disponibilidad y presencia generalizada. Nunca ha existido una mejor época para ser un experto en datos.