**¿Qué es un data scientist? Diferencias y similitudes entre data science y data analysis**

La ciencia de datos (data science) es una ciencia interdisciplinar que se nutre y utiliza múltiples áreas de conocimiento como la estadística, la programación e incluso otras ramas de la informática en combinación con la información que recibe del entorno de interés (una consultoría, un laboratorio que trabaja con datos biológicos, una industria cárnica, etc.) para entender, procesar, analizar y extraer conclusiones sobre los datos recogidos de dicho entorno. El científico de datos (data scientist), básicamente, se encarga de manejar y crear código utilizando conocimientos de informática y matemáticas para no solo analizar esos datos, sino llegar a sus propias conclusiones y sugerir y contestar a nuevas preguntas sobre el objeto de interés.

Un analista de datos (data analyst) también se encarga de procesar y analizar información para responder a preguntas a partir de un proceso de selección y análisis de datos. Sin embargo, las diferencias son claras entre el trabajo que realiza un data analyst y el trabajo que realiza un data scientist. Es más, en muchas empresas suelen convivir ambos tipos de especialistas que tienen focalizaciones distintas. Vemos las principales diferencias:

* En primer lugar, el data analyst normalmente solo responde a las preguntas que se le ofrecen en ese entorno de trabajo. Es decir, procesa y analiza los datos para obtener respuestas a las preguntas concretas que la empresa le encarga. Sin embargo, el data scientist va más allá y es capaz de entender la información de manera no tanto descriptiva sino para ser capaz de derivar nuevas preguntas a futuro.
* Por otro lado, en relación a lo anterior, el data analyst se centra más en lo que respecta a análisis descriptivo, visualización de datos y la identificación de patrones o tendencias para responder a esas preguntas (por ejemplo, hacer regresiones lineales o logísticas o un ANOVA/MANOVA de forma detallada para ver patrones, tendencias o interacciones entre datos). En cambio, el data scientist, debido a su enfoque más amplio, se centra más en análisis predictivo, modelización de datos, machine learning y sobre todo, en la generación de paquetes y software para ser usado a posteriori de ser necesario, reutilizando la información y los modelos para ser usados a futuro.
* Así, mientras que el data analyst no se sale tanto de su enfoque estadístico-matemático para ofrecer insights sobre los datos, el data scientist enfoca el problema desde una perspectiva más a futuro creando software que le pueda servir a la persona/empresa interesada, transformando los datos en modelos complejos para plantear otras preguntas, incluso por el propio data scientist.