

# Ciencia de datos

# Índice

**01** ¿Qué es la ciencia de datos?

**02** Importancia y utilidad de la ciencia de datos

**03** Herramientas y tecnologías

**04** Python en ciencia de datos



01

# ¿Qué es la ciencia de datos

# ¿Qué es la ciencia de datos?

Es un **campo interdisciplinario** que utiliza métodos científicos, procesos y sistemas para extraer conocimiento y entendimiento de datos estructurados y no estructurados.

- **Campo interdisciplinario:** Combina conocimientos de matemáticas, estadísticas, programación, y dominios específicos de aplicación.
- **Extracción de conocimiento:** Se enfoca en descubrir patrones, tendencias, y relaciones en los datos para generar insights.
- **Datos estructurados y no estructurados:** Incluye tanto datos organizados en tablas (como bases de datos) como datos no organizados (como texto, imágenes, audio).
- **Métodos científicos y procesos:** Aplica metodologías sistemáticas para recolectar, limpiar, analizar, modelar e interpretar datos.
- **Aplicaciones diversas:** Utilizado en una amplia gama de áreas, desde la medicina y las finanzas hasta la tecnología y el entretenimiento.

02

¿Para que se usa la  
ciencia de datos?

# Importancia y utilidad de la ciencia de datos

La **ciencia de datos** es de vital importancia en la actualidad debido a su capacidad para extraer conocimientos significativos de *grandes cantidades de datos*

- **Toma de decisiones informada:** Permite a empresas y organizaciones tomar decisiones basadas en datos en lugar de en suposiciones o intuiciones.
- **Optimización de procesos:** Ayuda a identificar áreas de mejora y optimización en diversos procesos, desde la producción hasta la logística.
- **Personalización y recomendaciones:** Facilita la creación de productos y servicios personalizados mediante el análisis de datos sobre el comportamiento y las preferencias de los usuarios.
- **Predicción y prevención:** Permite predecir tendencias futuras y anticipar problemas potenciales, como fraudes o fallas en sistemas.
- **Avances en la investigación y la ciencia:** Impulsa el progreso en diversas áreas, como la medicina, la genética, la climatología y la astrofísica, al permitir el análisis de grandes conjuntos de datos y la extracción de nuevos conocimientos.

03

# Herramientas y tecnologías

# Herramientas y tecnologías

En la **ciencia de datos**, se utilizan una variedad de **herramientas y tecnologías** para manipular, analizar y visualizar datos de manera efectiva

- **Python y R**: Son dos de los lenguajes de programación más populares para la ciencia de datos, gracias a sus numerosas bibliotecas y paquetes especializados.
- **Pandas**: Una biblioteca de Python que proporciona estructuras de datos flexibles y herramientas para el análisis de datos.
- **NumPy**: Ofrece soporte para arrays y matrices multidimensionales, junto con una amplia gama de funciones matemáticas para trabajar con estos datos.
- **Matplotlib y Seaborn**: Son bibliotecas de visualización de datos en Python que permiten crear gráficos y visualizaciones informativas.
- **Jupyter Notebook**: Una aplicación web que permite crear documentos interactivos que contienen código, visualizaciones y texto explicativo, lo que facilita la colaboración y la comunicación en proyectos de ciencia de datos.



04

# Python en ciencia de datos

# Python en ciencia de datos

**Python** se ha convertido en el lenguaje de programación preferido en la ciencia de datos debido a su versatilidad, facilidad de uso y robusta comunidad de desarrolladores.

Con una sintaxis clara y legible permite a los científicos de datos desarrollar rápidamente prototipos, experimentar con algoritmos y visualizar datos de manera efectiva.

Además cuenta con una amplia gama de bibliotecas especializadas, como Pandas, NumPy, Matplotlib y SciPy, que ofrecen herramientas poderosas para el análisis de datos, la manipulación de matrices, la visualización y la computación científica.

