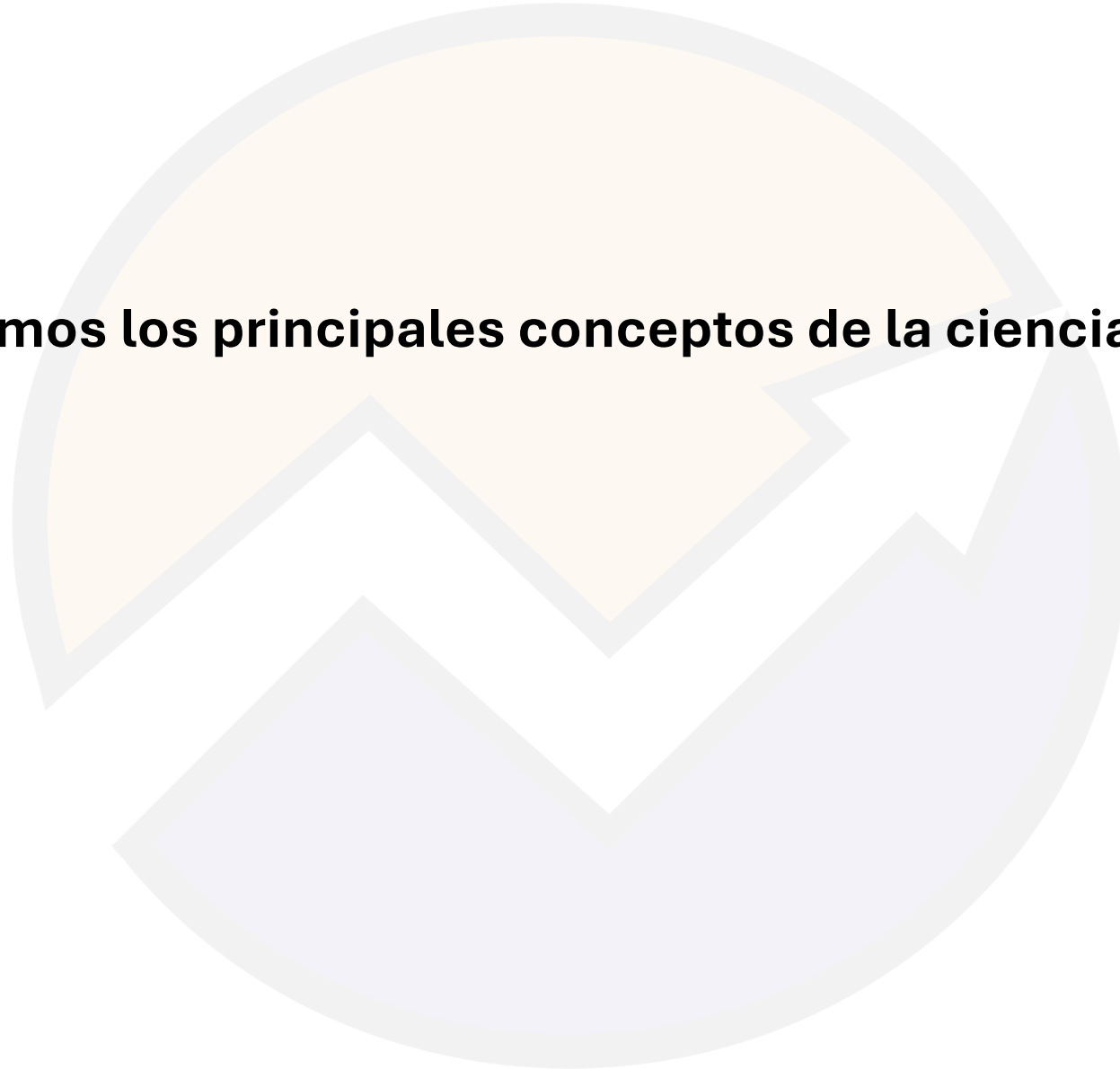


Mundo de la ciencia de datos

Aprenderemos los principales conceptos de la ciencia de datos.



¿Qué es la ciencia de datos?

La ciencia de datos es una disciplina que combina programación, estadísticas y conocimiento del negocio para **extraer valor a partir de los datos**.

Es un proceso que utiliza herramientas y técnicas para **recopilar, limpiar, analizar, visualizar y modelar datos** con el fin de responder preguntas, tomar decisiones o automatizar procesos.

La ciencia de datos incluye:

- Análisis exploratorio de datos
- Visualización de datos
- Modelos predictivos (machine learning)
- Automatización de reportes
- Toma de decisiones basada en datos



¿Qué NO es la ciencia de datos?

La ciencia de datos:

- No es solo hacer gráficos bonitos
- No es lo mismo que solo programar
- No es magia ni adivinación
- No es Excel con esteroides
- No es lo mismo que BI (Business Intelligence)



La ciencia de datos **no es una herramienta, sino una disciplina completa** que requiere pensamiento crítico, conocimiento del contexto y habilidades técnicas.

¿Por qué saber ciencia de datos?

Es importante saber ciencia de datos:

- Porque vivimos en un mundo gobernado por datos.
- Porque aumenta tus oportunidades laborales.
- Porque mejora la toma de decisiones.
- Porque te da superpoderes digitales.
- Porque aplica a cualquier industria.
- Porque no basta con tener datos, hay que saber usarlos.



¿Qué necesitamos para iniciar ciencia de datos?

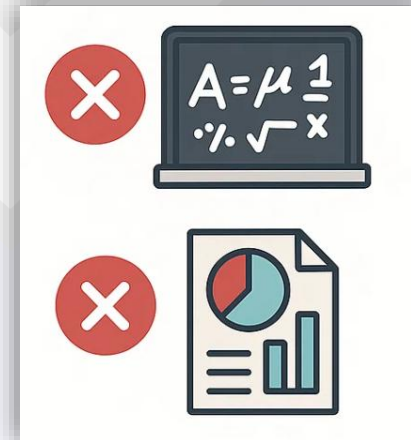
Para iniciar en ciencia de datos necesitamos:

- Mentalidad analítica.
- Conocimientos básicos de estadística.
- Aprender un lenguaje de programación.
- Dominar herramientas de análisis.
- Aprender haciendo.



Para iniciar en ciencia de datos NO necesitamos:

- Ser matemático.
- Saber álgebra lineal ni derivadas desde el día uno.
- Tener datos propios.



¿Qué es R y Python?

- R y Python son los dos lenguajes de programación más utilizados en el mundo de la ciencia de datos.
- Ambos son gratuitos, potentes y con enormes comunidades de usuarios.

| ¿Buscas...? | Usa... |
|-----------------------------------|------------------------|
| Análisis estadístico clásico | R |
| Machine learning y producción | Python |
| Facilidad para automatizar tareas | Python |
| Gráficos estadísticos potentes | R |
| Lo mejor de ambos mundos | Aprende los dos |

Aprenderemos ambos en este curso.

¿Cómo se relaciona la ciencia de datos con otras áreas?

La ciencia de datos no es un área aislada. Es una **herramienta transversal** que se integra en casi cualquier disciplina para potenciar decisiones, automatizar procesos o descubrir patrones.

Ciencia y salud

- Detección temprana de enfermedades.
- Predicción de brotes o riesgo.
- Análisis de eficacia de tratamientos.

Economía y finanzas

- Predicción de precios y demanda.
- Evaluación de riesgo crediticio.
- Análisis de inversiones.

Ingeniería y producción

- Mantenimiento predictivo.
- Optimización de procesos.
- Control de calidad automatizado.

Marketing y negocios

- Segmentación de clientes.
- Detección de abandono (churn).
- Campañas personalizadas.

Educación

- Detección de estudiantes en riesgo.
- Personalización del aprendizaje.
- Evaluación de resultados académicos.

Flujo típico de un proyecto de ciencia de datos

Aunque cada proyecto puede variar, la mayoría sigue un proceso estructurado que permite transformar datos en conocimiento útil y accionable.

1. Definición del problema

Antes de escribir una sola línea de código, debes entender:

- ¿Qué queremos resolver?
- ¿Cuál es el objetivo de negocio?
- ¿Qué decisiones se tomarán con este análisis?

2. Recolección de datos

Se obtienen los datos desde diferentes fuentes:

- Bases de datos internas
- APIs
- Archivos planos (Excel, CSV)
- Web scraping
- Encuestas

Flujo típico de un proyecto de ciencia de datos

Aunque cada proyecto puede variar, la mayoría sigue un proceso estructurado que permite transformar datos en conocimiento útil y accionable.

3. Limpieza y preparación de datos

Los datos vienen sucios: faltantes, duplicados, errores de formato, etc.

- Eliminación o imputación de valores nulos.
- Corrección de tipos de datos.
- Normalización de nombres, formatos, unidades.
- Creación de nuevas variables (features).

4. Análisis exploratorio (EDA)

Se exploran los datos para:

- Entender patrones, tendencias y outliers.
- Identificar relaciones entre variables.
- Visualizar y resumir información clave.

Flujo típico de un proyecto de ciencia de datos

Aunque cada proyecto puede variar, la mayoría sigue un proceso estructurado que permite transformar datos en conocimiento útil y accionable.

5. Modelado

Se entrena un modelo (si el objetivo lo requiere):

- Regresión.
- Clasificación.
- Clustering.
- Series de tiempo.
- Árboles, redes neuronales, etc.

6. Evaluación

Se mide el rendimiento del modelo con métricas:

- Clasificación: accuracy, precision, recall, F1, AUC.
- Regresión: MAE, RMSE, R^2 .

Flujo típico de un proyecto de ciencia de datos

Aunque cada proyecto puede variar, la mayoría sigue un proceso estructurado que permite transformar datos en conocimiento útil y accionable.

7. Comunicación de resultados

- Visualizaciones claras y explicativas.
- Tableros de control.
- Presentaciones con conclusiones para la toma de decisiones.

8. Despliegue y automatización

Si el modelo funciona, se puede automatizar:

- Reportes recurrentes.
- Integración en aplicaciones web.
- Automatización de decisiones.

En este curso veremos algunos ejemplos vitales de este proceso.

Conclusiones

La ciencia de datos **no es una herramienta, sino una disciplina completa** que requiere pensamiento crítico, conocimiento del contexto y habilidades técnicas.

Existen lenguajes de programación idóneos para ciencia de datos, como lo son Python y R.

La ciencia de datos se utiliza en múltiples industrias, por lo que es fundamental su dominio.

Existe un flujo normalizado de trabajo en ciencia de datos, pero puede ser ajustado.