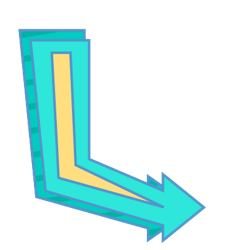
Definición y Conceptos Básicos

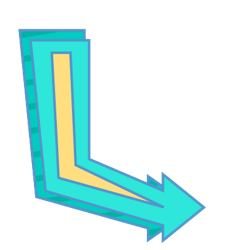
**Ciencia de Datos:**

La ciencia de datos es un campo interdisciplinario que utiliza métodos, procesos, algoritmos y sistemas para extraer conocimiento y obtener insights a partir de datos estructurados y no estructurados. Combina aspectos de estadística, informática y conocimiento del dominio para analizar e interpretar datos.



**Datos**:

Información sobre algo concreto que permite su conocimiento exacto o sirve para deducir consecuencias derivadas de un hecho, puede ser cuantitativa o cualitativa.

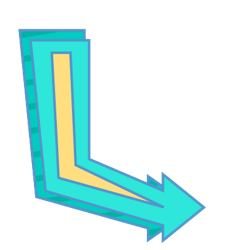


**Información:**

Datos procesados, organizados y presentados en un contexto significativo.

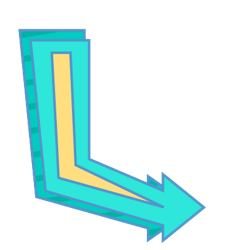
**Diferencia entre dato e información:**

Los datos son hechos sin procesar o sin contexto, mientras que la información es el resultado de la interpretación y organización de esos datos **EN DATA SCIENCE**.



**Datos:**

hechos o valores individuales, sin contexto ni significado inherente; pueden ser números, textos, imágenes, o cualquier otra representación de **información**; Son materia prima para construir la información



**Información:**

Es el resultado de procesar, organizar y contextualizar los datos; tiene un significado, una representación y propósito; se puede utilizar para tomar decisiones, comunicar conocimiento o resolver problemas

**Análisis de datos**:

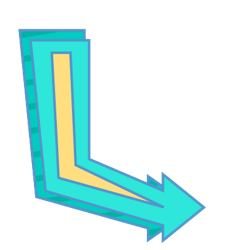
Aplicación de las estadísticas o análisis a algún tema/negocio/campo especifico. (búsqueda de conclusiones). Proceso de examinar, limpiar, transformar y modelar datos con el objetivo de descubrir información útil, extraer conclusiones y apoyar la toma de decisiones. Implica usar diversas herramientas, técnicas y metodologías para identificar patrones, tendencias y relaciones significativas dentro de conjuntos de datos.

**Big Data:**

Conjunto de datos extremadamente grandes y complejos que requieren técnicas avanzadas para su almacenamiento, procesamiento y análisis

**Machine Learning**:

Subcampo de la inteligencia artificial que se centra en el desarrollo de algoritmos que permiten a las máquinas aprender de los datos, patrones recurrentes en conjuntos de datos. Esos datos pueden ser números, palabras, imágenes, estadísticas.



**Deep Learning**:

Es una rama del machine learning, que utiliza redes neuronales artificiales con múltiples capas para analizar datos y aprender de ellos.

**Desarrollo de Software:**

Proceso de creación, diseño, implementación y mantenimiento de programa, aplicaciones o sistemas informáticos.Se utiliza ciencias de computacionales y código para resolver problemas de negocio o de la vida real, la diferencia entre programación y desarrollo de software es que la programación se centra en escribir código para realizar tareas específicas, mientras que el desarrollo de software abarca un proceso más amplio que incluye diseño, planificación, codificación, pruebas y mantenimiento del software

**Análisis Predictivo**:

Uso de modelos estadísticos y algorítmicos para predecir futuros eventos basados en datos históricos

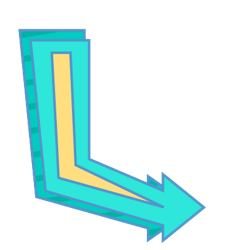
**Visualización de Datos**:

Procesos de representar la información y datos de forma gráfica, utilizando elementos como gráficos, diagramas, mapas y otros elementos visuales para facilitar la comprensión, exploración y comunicación de la información. Esto ayuda a identificar patrones, tendencias y relaciones en los datos, lo que a su vez permite tomar mejores decisiones y comunicar hallazgos de manera efectiva.

**Conceptos clave en Ciencia de datos**

**Recopilar datos:**

Es el punto de partida de cualquier proyecto de ciencia de datos. Aquí se obtienen los datos desde diversas fuentes. Esta etapa define la calidad y el potencial del análisis posterior



**Flexibilidad:**

Un científico de datos puede estudiar solamente imágenes de resonancia magnética, imágenes de automóviles, de tráfico, información del clima, lo que muestra la flexibilidad del rol del científico de datos.

**Limitantes**:

Limitantes procesamiento, por eso se debe evaluar cómo hacer o cómo se entrega más suave la información.

* Calidad de los datos (ruido, valores faltantes)
* Recursos computacionales
* Tiempo y presupuesto
* Interpretabilidad de los modelos
* Relación con análisis predictivo y prescriptivo: estas limitaciones afectan que tan complejos o precisos pueden ser los modelos predictivos o prescriptivos

**Limpieza de datos:**

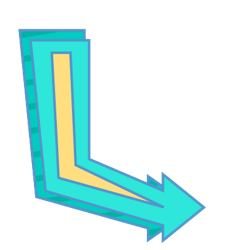
El trabajo que se realiza normalmente es 90% limpieza de datos y el 10% análisis de alto nivel.

**Importancia**:

La ciencia de datos es crucial en la era digital porque permite a las organizaciones tomar decisiones informadas basadas en datos. Se utiliza en diversas industrias como la salud, marketing, finanzas y la tecnología para mejorar procesos, optimizar recursos y descubrir nuevas oportunidades.

**Diferencia entre BI y DSc**: inteligencia de negocios (BI), hacia atrás descriptiva (pasado); ciencia de datos (DSc), es hacia adelante (futuro), es prescriptivo y predictivo.

* **Diagnostico descriptivo:**Quién, por qué, cuándo, quiénes, qué sucedió**BI**
* **Diagnóstico prescriptivo:** el por qué sucedió, métricas de causas. DSc
* **Análisis predictivo:** visión más probable del futuro o de una variable desconocida. BI
* **Análisis prescriptivo:**provee el mejor camino o mejor estrategia para alcanza un objetivo dado. DSc



|  |  |  |
| --- | --- | --- |
| **Aspecto** | **Business Intelligence (BI)** | **Ciencia de Datos (DSc)** |
| Enfoque | Descriptivo | Predictivo y prescriptivo |
| Tiempo | Pasado | Futuro |
| Herramientas | Dashboard, reportes | Modelos estadísticos, ML |
| Objetivo | Entender qué pasó | Predecir qué pasará y qué hacer |

**Tipos de Análisis**

* Descriptivo: ¿qué pasó? (BI)
* Diagnóstico: ¿Por qué pasó? (entre BI y DSc)
* Predictivo: ¿qué podría pasar? (DSc)
* Prescriptivo: ¿Qué deberíamos hacer? (DSc)

**Ciclo de vida:**

 Recopilar ->

Limpiar datos ->

Utilizar modelos para obtener información  ->

Conclusiones, Resultados y toma de decisiones

* **Descriptivo y diagnóstico**-> Durante el EDA (Exploratory Data Analysis)
* **Predictivo** -> Modelado
* **Prescriptivo** -> Implementación y toma de decisiones.

