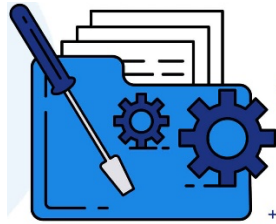


# ¿QUÉ ES UN CONJUNTO DE DATOS?



En el campo de la estadística y el análisis de información, un **conjunto de datos** se refiere a una recopilación organizada de observaciones o registros que contienen información sobre una o más variables de interés. Estos datos representan atributos o características medibles de un grupo de elementos u objetos de estudio, también conocidos como **unidades de observación**.

Cada observación dentro del conjunto de datos se asocia con **valores concretos para cada variable**, y su organización comúnmente se da en forma de tablas, donde las filas representan casos u observaciones individuales (como personas, productos, experimentos) y las columnas corresponden a las variables medidas (como edad, peso, rendimiento académico, etcétera).

Por ejemplo, al registrar las calificaciones de 10 alumnos, se obtiene el siguiente conjunto:  
**65, 70, 75, 80, 85, 85, 90, 90, 95, 100**

## Componentes básicos de un conjunto de datos

Un conjunto de datos típicamente está organizado en forma de tabla, en la cual:

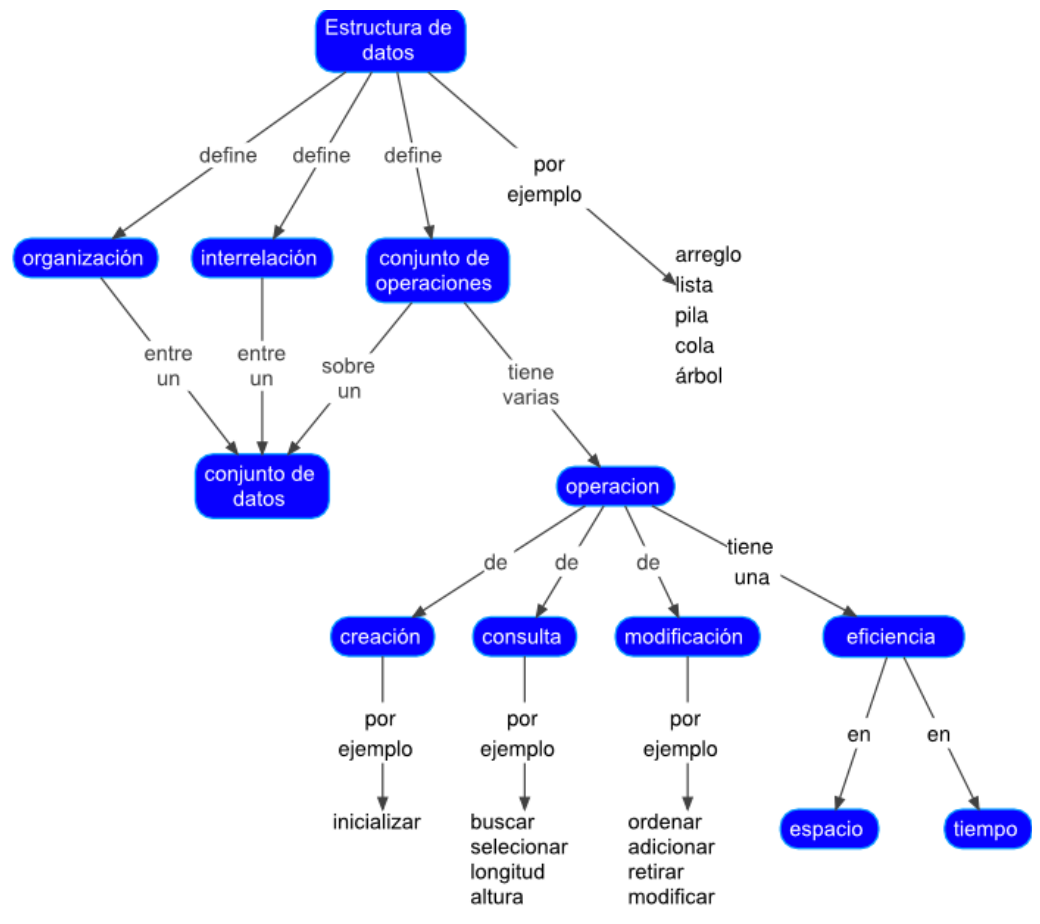
- Las **filas** representan las observaciones o casos (por ejemplo, cada persona encuestada).
- Las **columnas** corresponden a las variables medidas (por ejemplo, edad, género, ingreso).
- Las **celdas** contienen los valores específicos registrados para cada combinación de caso y variable.

	Disco HD	Memoria RAM	Pantalla	Precio Dólares
EJEMPLO DE ENTRENAMIENTO 1	80	4	14	\$1,499.00
EJEMPLO DE ENTRENAMIENTO 2	85	2	14	\$1,795.00
⋮	170	4	15	\$1,595.00
⋮	170	8	14	\$1,849.00
⋮	340	16	14	\$3,295.00
⋮	340	16	14	\$3,695.00
⋮	170	4	14	\$1,720.00
⋮	85	2	14	\$1,995.00
⋮	210	8	14	\$2,225.00
⋮	210	4	15	\$2,575.00
⋮	170	8	15	\$2,195.00
⋮	210	8	14	\$2,605.00
EJEMPLO DE ENTRENAMIENTO 14	130	4	14	\$2,045.00
⋮	245	8	14	\$2,295.00
⋮	212	8	14	\$2,699.00
EJEMPLO DE ENTRENAMIENTO N				

**CONJUNTO DE DATOS**

<https://share.google/images/wo9fokasLXuq4GoST>

### Tipos de conjuntos de datos según su estructura



<https://share.google/images/avFTm23Wo48o6moZI>

Para poder gestionar y analizar la información de manera eficaz, es necesario comprender que **no todos los conjuntos de datos tienen la misma forma ni el mismo nivel de organización**. En función de su estructura, los conjuntos de datos se pueden clasificar en:

### 1. Datos estructurados

Estos datos están organizados de manera precisa y sistemática, por lo general dentro de tablas con filas y columnas. Son fácilmente accesibles y analizables mediante herramientas estadísticas y bases de datos tradicionales.

**Ejemplos:**

- Bases de datos relacionales (SQL).
- Hojas de cálculo (por ejemplo, Excel).



#### ¿Cuándo usar datos agrupados?

- Cuando se tienen muchos datos (más de 20 o 30).
- Para facilitar cálculos estadísticos.
- Para construir histogramas o diagramas de frecuencia.

### 2. Datos no estructurados

Este tipo de datos carece de una organización formal que facilite su procesamiento con métodos estadísticos convencionales. A menudo, son de tipo textual o multimedia y requieren técnicas más avanzadas para su análisis.

**Ejemplos:**

- Documentos de texto (informes, correos electrónicos).
- Archivos de imagen, audio o video.
- Publicaciones en redes sociales.



#### ¿Cuándo usar datos no agrupados?

- Cuando se tienen pocos datos.
- Para mayor precisión en cálculos individuales.
- En estudios detallados o pequeños grupos (como en clases pequeñas).

### 3. Datos semi-estructurados

Estos datos se encuentran en un punto intermedio entre los estructurados y no estructurados. Aunque no están organizados en tablas formales, contienen etiquetas u organizadores que permiten cierto grado de procesamiento automatizado.

Ejemplos:

- Archivos **XML**.
- Formatos **JSON**, ampliamente utilizados en aplicaciones web.

#### Conjuntos de datos según su complejidad: El caso del *Big Data*

Con los avances tecnológicos, el volumen y la complejidad de los datos disponibles han aumentado de manera exponencial, dando origen a lo que se conoce como **grandes conjuntos de datos** o **big data**. Este tipo de datos no puede ser tratado eficazmente con herramientas estadísticas tradicionales debido a tres características clave, conocidas como las **3V del Big Data**:

- **Volumen**
- **Velocidad**
- **Variedad**



El trabajo con big data permite descubrir patrones complejos, comportamientos ocultos y relaciones significativas que antes eran difíciles de detectar. Este tipo de análisis se utiliza en campos como la **salud pública**, el **marketing digital**, la **inteligencia artificial**, la **ciencia del comportamiento** y las **finanzas**.



#### RECUERDA

Es importante identificar el tipo de dato, porque determina las herramientas estadísticas adecuadas para su análisis:

Tipo de dato	Medidas adecuadas
No agrupado	Media, mediana, moda, desviación típica
Agrupado	Media y varianza con fórmulas especiales
Cualitativo	Moda, porcentajes, gráficos de pastel
Cuantitativo	Todas las medidas estadísticas

**Referencia:**

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. Estados Unidos. O'Reilly Media.

Jonker, Alexandra; Gomstyn, Alice. (s.f.). *¿Cuál es la diferencia entre datos estructurados y no estructurados? IBM*. Recuperado de: <https://www.ibm.com/mx-es/think/topics/structured-vs-unstructured-data>

Calvo, Diego. (2017). *Tipos de datos: estructurados, semiestructurados y no estructurados*. Diego Calvo. *Análisis de datos*. Recuperado de: <https://www.diegocalvo.es/tipos-de-datos-estructurados-semiestructurados-y-no-estructurados>

Rodríguez Rodríguez, Alberto; Bernal Gamboa, Elizabeth. (2019) *Gestión de la Información Cuantitativa en las Universidades*. Universidad Nacional de Colombia. Recuperado de: [https://estadisticaun.github.io/L\\_Conceptual/2-1-los-datos.html](https://estadisticaun.github.io/L_Conceptual/2-1-los-datos.html)