

# Análisis de datos

## Nivel Básico – Explorador

### Misión 2

## Introducción a los Datasets en el Análisis de Datos

### *Contexto:*

En el análisis de datos, un dataset es una colección organizada de datos que puede ser utilizada para realizar análisis, construir modelos, y tomar decisiones basadas en datos. La calidad y la estructura de los datasets son cruciales para obtener resultados precisos y útiles. A continuación, qué son los datasets, los diferentes tipos de datasets que existen, cómo se estructuran y cómo se utilizan en el análisis de datos.

### *Objetivos:*

1. Comprender qué es un dataset y su importancia en el análisis de datos.
2. Identificar y diferenciar entre los tipos de datasets.
3. Aprender sobre la estructura de los datasets y las mejores prácticas para su manejo.
4. Explorar ejemplos prácticos de datasets y cómo prepararlos para el análisis.

## 1. Introducción a los Datasets

**1.1. Definición de Dataset** Un dataset es una colección de datos relacionados, organizados de manera estructurada. Los datasets suelen estar en formato de tablas, donde cada fila representa una observación o registro y cada columna representa una variable o característica.

**1.2. Importancia en el Análisis de Datos** Los datasets son la base del análisis de datos. La calidad del análisis depende de la calidad del dataset. Un dataset bien estructurado y limpio facilita la extracción de información valiosa y la construcción de modelos predictivos.

## 2. Tipos de Datasets

### 2.1. Datos Estructurados

- **Definición:** Datos que están organizados en un formato tabular, como hojas de cálculo o bases de datos relacionales.
- **Ejemplos:** Datos en hojas de cálculo de Excel, bases de datos SQL.

### 2.2. Datos No Estructurados

- **Definición:** Datos que no tienen una estructura predefinida, como textos, imágenes o videos.
- **Ejemplos:** Correos electrónicos, publicaciones en redes sociales, archivos de texto libre.

### 2.3. Datos Semi-Estructurados

- **Definición:** Datos que no se ajustan a un formato tabular rígido, pero tienen algún tipo de etiqueta o estructura.
- **Ejemplos:** Archivos JSON, XML, logs de servidor.

## 3. Estructura de un Dataset

### 3.1. Componentes Clave

- **Filas (Registros):** Cada fila representa una unidad de análisis, como una transacción o una persona.
- **Columnas (Variables):** Cada columna representa una característica o variable que describe los registros.
- **Metadatos:** Información adicional que describe el dataset, como la fuente de los datos, el formato y las fechas de captura.

### 3.2. Ejemplo de Estructura

Imaginemos un dataset de ventas:

Fecha	Producto	Cantidad	Precio Unitario	Total
2024-07-01	Widget A	10	15.00	150.00
2024-07-01	Widget B	5	25.00	125.00

- **Fecha:** Variable de fecha.
- **Producto:** Variable categórica.
- **Cantidad:** Variable numérica.
- **Precio Unitario:** Variable numérica.
- **Total:** Variable calculada (Cantidad \* Precio Unitario).

## 4. Manejo y Preparación de Datasets

### 4.1. Limpieza de Datos

- **Eliminar Duplicados:** Asegurarse de que no haya registros repetidos.
- **Manejo de Valores Faltantes:** Imputar valores faltantes o eliminar registros con datos incompletos.
- **Corrección de Errores:** Arreglar errores en los datos, como valores fuera de rango o errores tipográficos.

### 4.2. Transformación de Datos

- **Normalización:** Escalar los datos para que estén en un rango común.
- **Codificación:** Convertir variables categóricas en un formato numérico.
- **Agrupación y Resumen:** Agrupar datos para obtener resúmenes o estadísticas.

**4.3. Ejemplo de Limpieza y Preparación** Supongamos que tenemos un dataset de ventas con valores faltantes en la columna de "Cantidad". Podemos decidir imputar estos valores con la media de la columna o eliminar las filas con valores faltantes.

## 5. Ejemplos Prácticos y Actividades

### 5.1. Actividad 1: Exploración de un Dataset

- **Objetivo:** Familiarizarse con un dataset real.
- **Instrucciones:** Proporcione un dataset simple (como el de ventas mencionado) y pida a los alumnos que exploren los datos, identifiquen problemas y propongan soluciones.

### 5.2. Actividad 2: Limpieza y Transformación

- **Objetivo:** Aplicar técnicas de limpieza y transformación.
- **Instrucciones:** Pida a los alumnos que realicen limpieza y transformación en un dataset proporcionado. Esto puede incluir eliminar duplicados, manejar valores faltantes y normalizar los datos.

### 5.3. Actividad 3: Análisis y Visualización

- **Objetivo:** Analizar y visualizar datos.
- **Instrucciones:** Utilizando el dataset limpio, pida a los alumnos que realicen un análisis simple y creen visualizaciones (gráficos de barras, histogramas) para interpretar los datos.