

La **tarea de revisión de datos** es un paso fundamental para *entender y preparar un conjunto de datos* antes de usarlo en cualquier análisis o modelo. Este proceso normalmente incluye varias etapas clave:

1. Exploración Inicial:

- Revisión de las dimensiones del conjunto de datos: número de filas (observaciones) y columnas (variables).
- Identificación de los tipos de datos presentes (numéricos, categóricos, texto, fechas, etc.).
- Observación de la distribución de los datos y las primeras muestras para captar su estructura.

2. Detección de Errores o Valores Faltantes:

- Identificación de datos inconsistentes, errores tipográficos, valores duplicados o anómalos.
- Verificación de valores nulos o ausentes y su porcentaje en cada variable.

3. Análisis Estadístico Básico:

- Cálculo de medidas descriptivas como la media, la mediana, la desviación estándar, los valores mínimos y máximos.
- Generación de distribuciones de frecuencia para datos categóricos.

4. Análisis de Relación entre Variables:

- Detección de correlaciones entre variables numéricas.
- Exploración de relaciones entre variables categóricas y numéricas mediante gráficos o tablas cruzadas.

5. Evaluación de Calidad de los Datos:

- Identificación de columnas irrelevantes o redundantes que podrían descartarse.
- Evaluación de la completitud, consistencia y precisión de los datos.

6. Documentación:

- Registro de las observaciones encontradas y cualquier cambio realizado durante la revisión.
- Anotación de características importantes del conjunto de datos para su referencia futura.

Este proceso no solo ayuda a entender el conjunto de datos, sino también a identificar problemas que pueden influir en los análisis posteriores. Es como preparar un lienzo limpio antes de pintar una obra.