**Análisis de datos**

Es el proceso de limpiar, inspeccionar, transformar y modelar datos con el objetivo de descubrir información útil para conclusiones o toma de decisiones a través de grandes cantidades de información.

**Fases del análisis de datos.**

**Recolección de datos:** Obtener datos de fuentes como BD, encuestas, Excel. Los datos pueden ser estructurados en tablas, o no se estructurados (imágenes, texto).

**Limpieza de datos**: Pueden contener errores, valores duplicados, faltantes, para asegurar que el análisis sea preciso.

**Exploración de datos:** Se usa la estadística descriptiva para comprender la estructura y las propiedades de los datos. Visualizar datos, patrones de repetición, anomalías, etcétera.

**Transformación y modelado**: Se trasforman y se preparan para ser analizados. Se aplica el Machine learning o modelos estadísticos para predicciones o clasificaciones.

**Herramientas para el análisis de datos.**

**Numpy** es fundamental para la computación científica, especialmente en el manejo de arrays y operaciones numéricas. Proporciona soporte al crear y manipular matrices (Arr) multidimensionales, y crear funciones matemáticas sobre los mismos.

La librería de numpy sirve para el manejo de matrices bidimensionales y multidimensionales, se toman de referencia los datos de matrices creadas manualmente, con comandos para crear matrices con ceros, unos, números aleatorios, números determinados en cierto rango, crear matrices simples de una sola dimensión con N numero de elementos y se transforman a bidimensionales con esos N elementos.

**Pandas** sirve para la manipulación y el análisis de datos en estructuras tabulades como hojas de cálculo o bases de datos. Permite acceder a datos mediante índices o nombres para filas y columnas. Ofrece métodos para reordenar, dividir y combinar conjuntos de datos.

Proporciona herramientas flexibles y eficientes para trabajar con datos estructurados, como tablas y series temporales. Algunas de las características clave de Pandas incluyen:

**Estructuras de datos principales.**

* DataFrame: Una estructura de datos bidimensional parecida a una tabla de Excel, hojas de cálculo, con filas, columnas etiquetadas.
* Series: Una estructura unidimensional, similar a una columna o un array.

**Funciones comunes de Pandas.**

* Lectura y escritura de datos desde diferentes formatos (**CSV, Excel, SQL, JSON, entre otros**).
* Filtrado, selección y modificación de datos.
* Agrupación, agregación y transformación de datos.
* Limpieza de datos como la gestión de valores nulos.
* Análisis y visualización de datos.