



M3: Obtención y preparación de datos

1. **DataFrame:** Estructura bidimensional de Pandas que permite almacenar y manipular datos en forma de tabla con filas y columnas etiquetadas.
2. **Serie:** Estructura unidimensional de Pandas que almacena una secuencia de datos con un índice asociado a cada elemento.
3. **Imputación:** Técnica utilizada para reemplazar valores perdidos en un conjunto de datos, utilizando estadísticas como la media, mediana o modelos predictivos.
4. **Outlier:** Valor atípico que se desvía significativamente del resto de los datos y puede afectar la calidad del análisis estadístico.
5. **Indexación jerárquica:** Método en Pandas para estructurar datos con múltiples niveles de índice, facilitando la segmentación y el acceso eficiente.
6. **Data Wrangling:** Proceso de limpieza, transformación y enriquecimiento de datos para convertir información cruda en conjuntos listos para el análisis.
7. **Melt:** Técnica de Pandas para transformar un DataFrame de formato ancho a formato largo, facilitando la normalización de datos.
8. **Pivot:** Proceso que permite reorganizar un DataFrame para convertir los valores únicos de una columna en encabezados de nuevas columnas.
9. **Merge:** Operación en Pandas que combina dos DataFrames en base a una o varias claves comunes, similar a las uniones de SQL.
10. **Concatenación:** Técnica que permite unir múltiples DataFrames a lo largo de filas o columnas para consolidar la información.
11. **NaN:** Valor especial en Pandas que representa datos faltantes o valores no disponibles en un conjunto de datos.

- 12.Z-score:** Métrica estadística que indica cuántas desviaciones estándar se encuentra un dato respecto a la media de su conjunto.
- 13.IQR (Rango Intercuartil):** Medida estadística que define el rango entre el primer y tercer cuartil, utilizada para detectar outliers.
- 14.Read_csv:** Función de Pandas que permite leer un archivo CSV y convertirlo en un DataFrame para su análisis y manipulación.
- 15.To_excel:** Método de Pandas que permite exportar un DataFrame a un archivo Excel para su almacenamiento o distribución.
- 16.Arreglo:** Estructura de datos de NumPy que permite almacenar múltiples elementos numéricos en un solo contenedor eficiente.
- 17.Reshape:** Función de NumPy utilizada para cambiar la forma o dimensión de un arreglo sin alterar sus datos.
- 18.Broadcasting:** Mecanismo de NumPy que permite realizar operaciones aritméticas entre arreglos de diferentes dimensiones de manera eficiente.
- 19.Duplicados:** Registros repetidos en un conjunto de datos que deben ser identificados y eliminados para evitar sesgos en el análisis.
- 20.Apply:** Método de Pandas que permite aplicar funciones personalizadas a las filas o columnas de un DataFrame para transformar los datos.