

Análisis de datos

Nivel Básico – Explorador

Misión 2

Exploración inicial de datos: Pandas

Operación sobre los datos en Pandas:

Una de las piezas esenciales de NumPy es la capacidad de realizar operaciones vectorizadas, tanto con aritmética básica (suma, resta, multiplicación, etc.) como con operaciones más sofisticadas (funciones trigonométricas, funciones exponenciales y logarítmicas, etc.). Pandas hereda gran parte de esta funcionalidad de NumPy.

Sin embargo, Pandas incluye un par de elementos útiles: para operaciones unitarias como funciones de negación y trigonométricas, estas funciones universales conservarán etiquetas de índice y columna en la salida, y para operaciones binarias como suma y multiplicación, Pandas alineará automáticamente los índices.

Glosario:

- **Operación unaria:** Es un tipo de operación matemática que necesita un operador y solo un argumento para ser realizada y obtener un resultado.
- **Operación binaria:** Es un tipo de operación matemática que necesita un operador y dos argumentos para obtener un resultado. Las operaciones matemáticas básicas son operaciones binarias.
- **Operación vectorizada:** Es una operación matemática que implica vectores y aborda un gran conjunto de operaciones, desde las operaciones básicas hasta las más complejas.

- **Índice:** Es una relación ordenada que define la posición de un dato y permite acceder a él de manera sencilla.
- **Archivo CSV:** Es un tipo de archivo que almacena la información en valores separados por comas. Normalmente se usa para el análisis de datos.
- **Datos desconocidos o nulos:** Son los tipos de datos nulos None o desconocidos NaN que se encuentran en una serie. Son indeseados en el análisis de datos, por lo cual es mejor eliminarlos.
- **Filtrado:** Es una técnica en el análisis de datos que permite obtener los valores que se desean de la base de datos bajo una condición impuesta.
- **Método in place:** Es un tipo de método en el análisis de datos que modifica el dataframe de manera directa.