



DigitalHouse >
Coding School

DATA SCIENCE

UNIDAD 1

Introducción a Pandas

INTRODUCCIÓN A PANDAS



1

Aprender las bases de pandas: para qué sirve, cuáles son sus estructuras de datos y aplicaciones principales.

2

Aplicar los conocimientos adquiridos de pandas en prácticas guiadas y labs.

- Fortalezas y limitaciones de **Numpy: debatir en clase**

- **Colección de estructuras de datos y funciones** que facilitan el trabajo con datos estructurados.
- Nombre derivado de "**Panel Data System**" (término econométrico para datasets multidimensionales).

- Construido **en base a Numpy**, inicialmente por Wes McKinney.
- Ofrece más versatilidad que Numpy: podemos trabajar con **datos de diferentes tipos** y podemos identificar las filas y columnas **usando etiquetas**, en lugar de solamente enteros.
- Combina la alta performance de las **operaciones sobre arrays de NumPy con la flexibilidad en la manipulación de datos** de un spreadsheet o una base de datos relacional.

- Provee **funcionalidades de indexación avanzadas** para facilitar la manipulación, agregación y selección de partes de un dataset.
- Provee **operaciones de agrupación por columnas, filtros y sumalizaciones.**
- Estas características hacen de pandas una librería **indispensable para las tareas de limpieza y preprocesamiento de datos.**

- Veremos los siguientes objetos:
 - Series
 - DataFrames
 - Index

	Animales
0	Perro
1	Oso
2	Jirafa
3	Tigre
4	Serpiente
5	Ratón

Una Series es un objeto similar a un **vector uni-dimensional**.



The diagram illustrates a Pandas Series as a table with two columns. The first column is labeled 'Index' and contains numerical values from 0 to 5. The second column is labeled 'Animales' and contains the names of animals: Perro, Oso, Jirafa, Tigre, Serpiente, and Ratón. A red arrow points from the 'Animales' header to the word 'Name'. A red bracket on the right side of the table groups the animal names under the label 'Values'.

	Animales
0	Perro
1	Oso
2	Jirafa
3	Tigre
4	Serpiente
5	Ratón

Contiene **un array de valores y un array asociado de etiquetas** de estos valores denominado como índice.

The diagram illustrates a Pandas Series as a table with two columns. The first column is labeled 'Index' and contains values from 0 to 5. The second column is labeled 'Animales' and contains the names of animals. A red arrow points from the 'Animales' header to the word 'Name'. A red bracket on the right side of the table groups the animal names under the label 'Values'.

	Animales
0	Perro
1	Oso
2	Jirafa
3	Tigre
4	Serpiente
5	Ratón

Una Serie también puede ser pensada como un **diccionario de tamaño fijo con sus claves ordenadas**.

The diagram shows a Pandas Series table with two columns: 'Index' and 'Animales'. The 'Index' column contains values from 0 to 5. The 'Animales' column contains the names of animals: Perro, Oso, Jirafa, Tigre, Serpiente, and Ratón. A red bracket on the left side of the table is labeled 'Index'. A red bracket on the right side of the table is labeled 'Values'. A red arrow points from the 'Animales' header to the word 'Name'.

	Animales
0	Perro
1	Oso
2	Jirafa
3	Tigre
4	Serpiente
5	Ratón

Al igual que los arrays de NumPy, permiten pasar una **lista de valores con índices** para seleccionar un subconjunto de valores.

Eje 1 (columnas) →

Eje 0 (filas) ↓

	Animales	Dueños
0	Perro	Juan
1	Oso	Pedro
2	Jirafa	Cristian
3	Tigre	Esteban
4	Serpiente	Pablo
5	Ratón	Claudio

Representa una **estructura de datos tabular** que contiene una **colección de columnas**, cada una de las cuales tiene un tipo determinado (number, string, boolean, etc.).

Eje 1 (columnas) →

Eje 0 (filas) ↓

	Animales	Dueños
0	Perro	Juan
1	Oso	Pedro
2	Jirafa	Cristian
3	Tigre	Esteban
4	Serpiente	Pablo
5	Ratón	Claudio

El DataFrame tiene **índices de columnas y filas**.

CONCLUSIONES



- En el día a día de un data scientist la limpieza, preparado y normalización de los datos con los que trabaja es la tarea que más tiempo insume.
- Pandas es la librería principal al momento de tener que realizar estas tareas.
- Los DataFrame de Pandas son la herramienta fundamental que vamos a utilizar a lo largo del curso y de nuestro día de trabajo a partir de ahora.