

Librerías para ciencia de datos – Pandas

Las herramientas del científico de datos

Juan Manuel Moreno — jmmoreno@profesorescol.imf.com



ÍNDICE

1. Objetivos unidad 2
2. Pandas
3. Anexo Join

01

Objetivos unidad 2

1.– Objetivos Unidad 2

- Conocer los principales tipos de archivos con los que podemos trabajar en Python.
- Saber realizar operaciones de lectura y escritura de archivos.
- Aprender a utilizar arrays uni–dimensionales y multi–dimensionales.
- Diferenciar las diferentes dimensiones de un array.
- Conocer las principales funciones de cálculo numérico, vectorial y matricial sobre arrays con Numpy.
- Comprender qué es un objeto Series y cuál es su relación dentro de los Dataframes.
- Conocer las funcionalidades y transformaciones fundamentales que podemos realizar en dataframes.



02

Pandas

2.1.- ¿Qué es Pandas?

- Se trata de una librería dedicada al procesamiento de datos en forma de matrices o tablas de datos (PANel DATA).
- Una de las librerías por excelencia para realizar ciencia de datos.
- Estructuras de datos con las que se trabaja: Series y Dataframes.
- Una sucesión de series compone un Dataframe
- Puede ser necesaria su instalación `pip install pandas`
- Se utiliza comúnmente con el alias `pd`
- Para importar la librería se emplea `import pandas as pd`



2.2.- Series

- Son la unidad mínima como estructura de datos para poder constituir un dataframe.
- Cualquier columna de un dataframe es un objeto series.
- En una serie todos sus elementos se distribuyen de forma clave : valor
- Un dataframe por lo tanto es una sucesión de series que comparten el mismo índice

```
criptos = pd.Series(  
    [15613.77, 496.29, 0.52, 0.83, 242.03],  
    index= ['Bitcoin', 'Ethereum', 'XRP', 'Tether', 'BTC Cash']  
)
```

criptos

Bitcoin	15613.77
Ethereum	496.29
XRP	0.52
Tether	0.83
BTC Cash	242.03

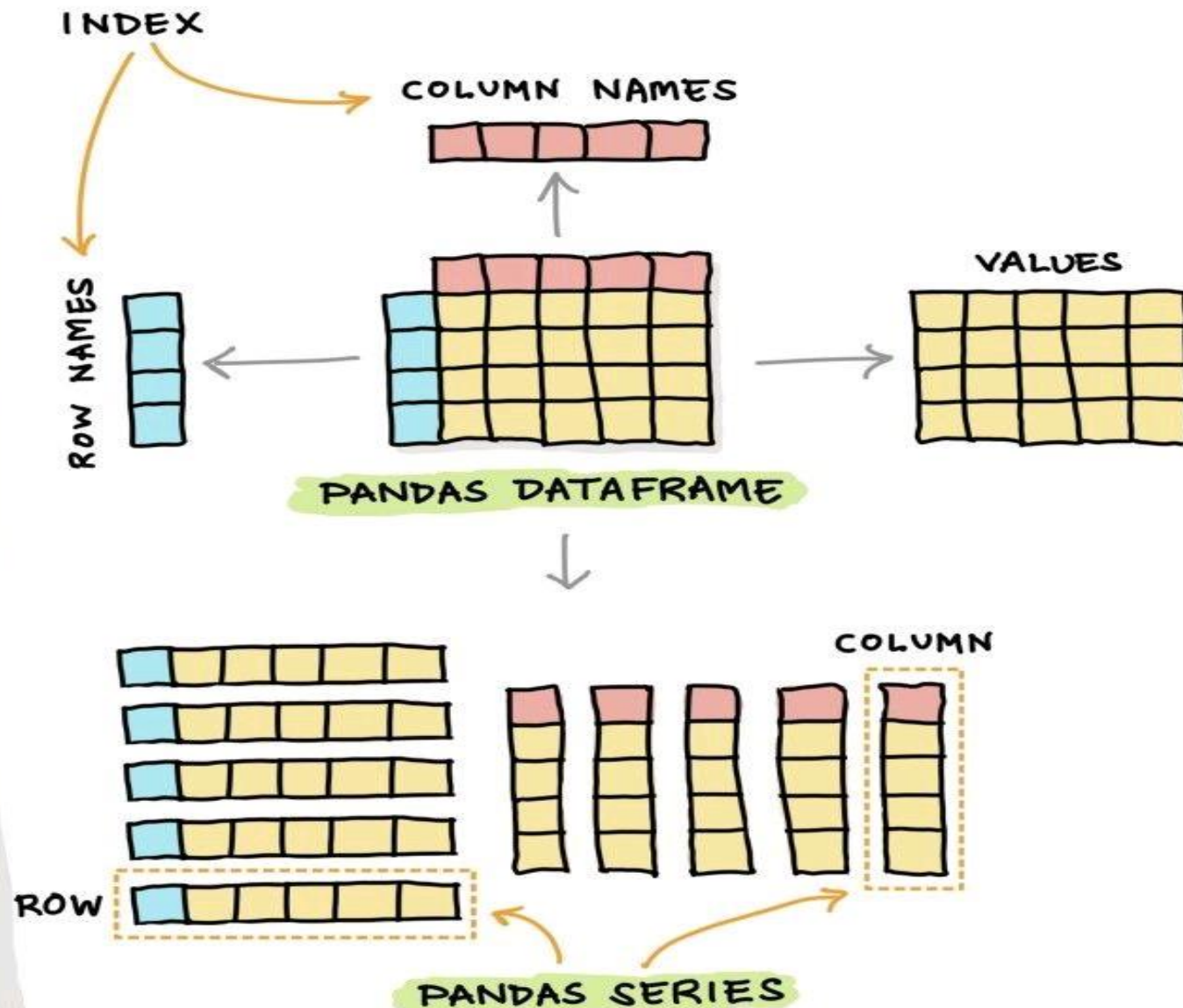
dtype: float64

2.3.– Dataframes

- Se construyen mediante la librería `pandas.DataFrame()`, son por lo tanto, una sucesión de series, en esta sesión veremos...
 - Estructuras de datos como dataframes.
 - Csv como dataframes.
 - Series y columnas de un dataframe.
 - Selección de filas y columnas de un dataframe (indexación)
 - Índices de un dataframe
 - Funciones básicas sobre dataframes.
 - Agrupación de dataframes.
 - Merge (unión) de dataframes
 - Gestión de columnas (agregación, borrado y transformación).
 - Gestión de datos nulos

2.- Pandas

ANATOMY OF PANDAS DATA STRUCTURES



Seguimiento práctico del contenido

A partir de aquí, vamos a ver las principales funciones, operaciones y curiosidades que podemos realizar con Pandas.

2_2_Pandas.ipynb

2_2_1_Ejercicio_complementario_Pandas_Numpy.ipynb



03

Anexo Join

Left Join

1	x1	1	y1
2	x2	2	y2
3	x3	4	y4
		2	y5

Right Join

1	x1	1	y1
2	x2	2	y2
3	x3	4	y4

Inner Join

1	x1	1	y1
2	x2	2	y2
3	x3	4	y4

Full Join

1	x1	1	y1
2	x2	2	y2
3	x3	4	y4

Full Join

1	x1	1	y1
2	x2	2	y2
3	x3	4	y4

IMF

Smart Education