



# CHEATSHEET PANDAS – ESTRUCTURA DE DATOS

## SET UP

- **Instalación:** `conda install pandas`
- **Importación:** `import pandas as pd`
- **Para el autocompletar:** `%config IPCompleter.greedy=True`

## ESTRUCTURA DE UN OBJETO SERIES

Axis 0 / index



## ESTRUCTURA DE UN OBJETO DATAFRAME

Axis 1 / columns

Axis 0 / index

Nombres de las variables

	id	Funded Date	Funded Amount	Country	Country Code	Loan Amount	Paid Date	Paid Amount	Activity	Sector	Delinquent	Name
0	84	2005-03-31T06:27:55Z	500	Uganda	UG	500	2005-12-13T12:00:40Z	500.0	Butcher Shop	Food	False	Justin
1	85	2005-03-31T06:27:55Z	500	Uganda	UG	500	2005-12-13T12:04:15Z	500.0	Food Production/Sales	Food	False	Geoffrey
2	86	2005-03-31T06:27:55Z	500	Uganda	UG	500	2005-12-13T12:06:56Z	500.0	Animal Sales	Agriculture	False	Apollo
...	...	...	...	...	...	...	...	...	...	...	...	...
5143	5257	2007-06-02T03:57:56Z	1000	Ghana	GH	1000	2008-03-13T10:01:32Z	1000.0	Construction	Construction	False	Abba
5144	5258	2007-03-24T06:01:10Z	1100	Ghana	GH	1100	2007-12-26T10:15:55Z	1100.0	Construction	Construction	False	Manas
5145	5259	2007-06-05T04:58:45Z	800	Ghana	GH	800	2008-03-13T10:01:33Z	800.0	Food Market	Food	False	Deasy

5146 rows x 13 columns

Diagram illustrating the structure of a Pandas DataFrame:

- Axis 1 / columns:** A horizontal axis pointing to the right.
- Axis 0 / index:** A vertical axis pointing downwards.
- Nombres de las variables:** Labels for the columns (id, Funded Date, Funded Amount, Country, Country Code, Loan Amount, Paid Date, Paid Amount, Activity, Sector, Delinquent, Name).
- Valores:** The data values within the DataFrame cells.
- Datos truncados:** Indicated by ellipses in the index and column labels.

# CHEATSHEET PANDAS – ESTRUCTURA DE DATOS

## TIPOS DE DATOS

- **float:** es el float de Numpy, que soporta datos nulos
- **int:** es el int de Numpy, que NO soporta datos nulos
- **Int64:** es un int que introduce Pandas para soportar datos nulos
- **object:** es el object de Numpy, que se usa como texto para meter cualquier cosa
- **category:** es un tipo que introduce Pandas específicamente para variables categóricas
- **bool:** es el booleano de Numpy, que NO soporta datos nulos
- **boolean:** es el booleano que introduce Pandas para soportar datos nulos
- **datetime64:** es el tipo de Numpy para las fechas, que NO soporta datos nulos

## CONVERSIONES DE ESTRUCTURAS DE DATOS

- **Crear un Series a partir de una lista:** `pd.Series(['a','b','c'])`
- **Crear un Series a partir de un diccionario:** `pd.Series({'a': 1, 'b': 2, 'c': 3})`
- **Crear un Series a partir de un array de numpy:** `pd.Series(np.array([1,2,3]))`
- **Pasar de Series a Dataframe:** `serie.to_frame()`
- **Crear un Dataframe a partir de una lista con sus componentes:**  
`pd.DataFrame(data = datos_para_df, columns=columnas_para_df, index = indice_para_df)`
- **Crear un Dataframe a partir de un diccionario:**  
`pd.DataFrame({'letras': ['a','b','c'], 'numeros': [1,2,3]})`
- **Crear un Dataframe a partir de un array de 2 dimensiones:**  
`pd.DataFrame(vector_2d, index = ['fila1','fila2'], columns = ['col1','col2','col3'])`
- **Pasar de Dataframe a Series (notación de punto):** `df.var1`
- **Pasar de Dataframe a Series (notación de corchetes):** `df['var1']`