

# Data Analysis con Python

Juan Felipe Acevedo Pérez  
Monitor Junior

Curso libre:  
Data Analysis con  
Python

Monitor encargado:  
Juan Felipe Acevedo  
Pérez

**Correo:** [uniic\\_bog@unal.edu.co](mailto:uniic_bog@unal.edu.co)

**Tel:** 3165000 **Ext:** 12301

# 5

## Sesión 5

*Análisis de datos -Visualización*

Correo: [uniic\\_bog@unal.edu.co](mailto:uniic_bog@unal.edu.co)

Teléfono: 3165000 ext 12301

# Contenido



**Gestión de datos  
(unir y combinar)**



**Reformar y  
pivotar**



**Visualización de  
datos**

# Gestión de datos

En muchas aplicaciones, los datos pueden distribuirse entre varios archivos o bases de datos o pueden organizarse en una forma que no es fácil de analizar.

La primera parte de la sesión se centra en las herramientas para ayudar a **combinar, unir y reorganizar datos**.

## Pandas DataFrame.merge()

The Left DataFrame

	Key	B
0	Key_0	145
1	Key_1	2373
2	Key_2	415
3	Key_3	2946

The Inner join

	key	B_x	A	B_y
0	Key_0	145	113	991.03
1	Key_1	2373	2342	993.13
2	Key_4	415	2234	995.44

Output

The Right DataFrame

	Key	A	B
0	Key_0	113	991.03
1	Key_1	2342	993.13
2	Key_2	4567	983.12
3	Key_3	2563	936.45
4	Key_4	2234	995.44
5	Key_5	71218	999.99



www.educba.com

Correo: uniic\_bog@unal.edu.co

Teléfono: 3165000 ext 12301

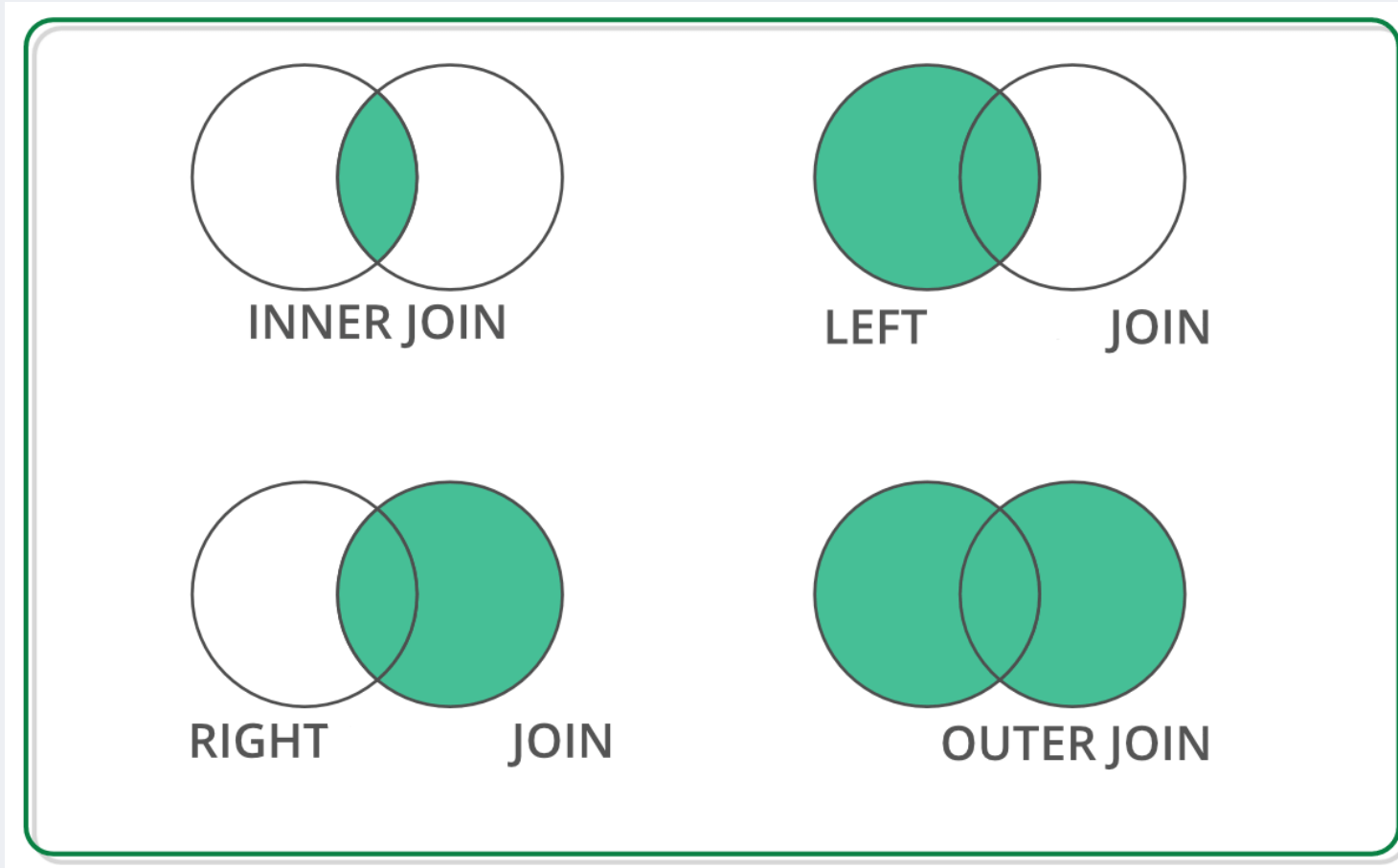


# Combinar y fusionar DFs

Los datos contenidos en los objetos pandas se pueden combinar de varias maneras:

- **pandas.merge()** conecta filas en DataFrames en función de una o más claves. Esto resultará familiar para los usuarios de SQL u otras bases de datos relacionales, ya que implementa operaciones de combinación de bases de datos.
- **pandas.concat()** concatena o “apila” objetos a lo largo de un eje.
- El método **combine\_first** permite unir datos superpuestos para completar los valores que faltan en un objeto con valores de otro

# Merge




# Reformar y pivotar

Hay una serie de operaciones básicas para reorganizar datos tabulares. Estas se conocen alternativamente como operaciones de **cambio de forma o de pivote**.

Su objetivo es remodelar los datos (produzca una tabla "pivot") en función de los valores de las columnas. Utilizando valores únicos de índices/columnas especificados para formar ejes del DataFrame resultante.

Dataset						
Name	Gender	Age				
John	Male	45				
Sammy	Female	6				
Stephan	Male	4				
Joe	Female	36				
Emily	Female	12				
Tom	Male	43				



Pivot Table			
Gender	%Gender	Age Group	Count
Male	50%	>18 years	2
		<18 years	1
Female	50%	>18 years	1
		<18 years	2

# Visualización de datos



- Hacer visualizaciones informativas (a veces llamadas gráficos) es una de las tareas más importantes en el análisis de datos. Puede ser parte del proceso **exploratorio**, por ejemplo, para ayudar a identificar valores atípicos o transformaciones de datos necesarias, o como una forma de generar ideas para **modelos**.
- **matplotlib** es un paquete de gráficos de escritorio diseñado para crear (en su mayoría bidimensionales) gráficos con calidad de publicación. El proyecto fue iniciado por John Hunter en 2002 para habilitar una interfaz de trazado similar a MATLAB en Python.