

DigitalHouse >
Coding School

DATA SCIENCE

MÓDULO 2

Funciones de agrupación
y agregación con Pandas:

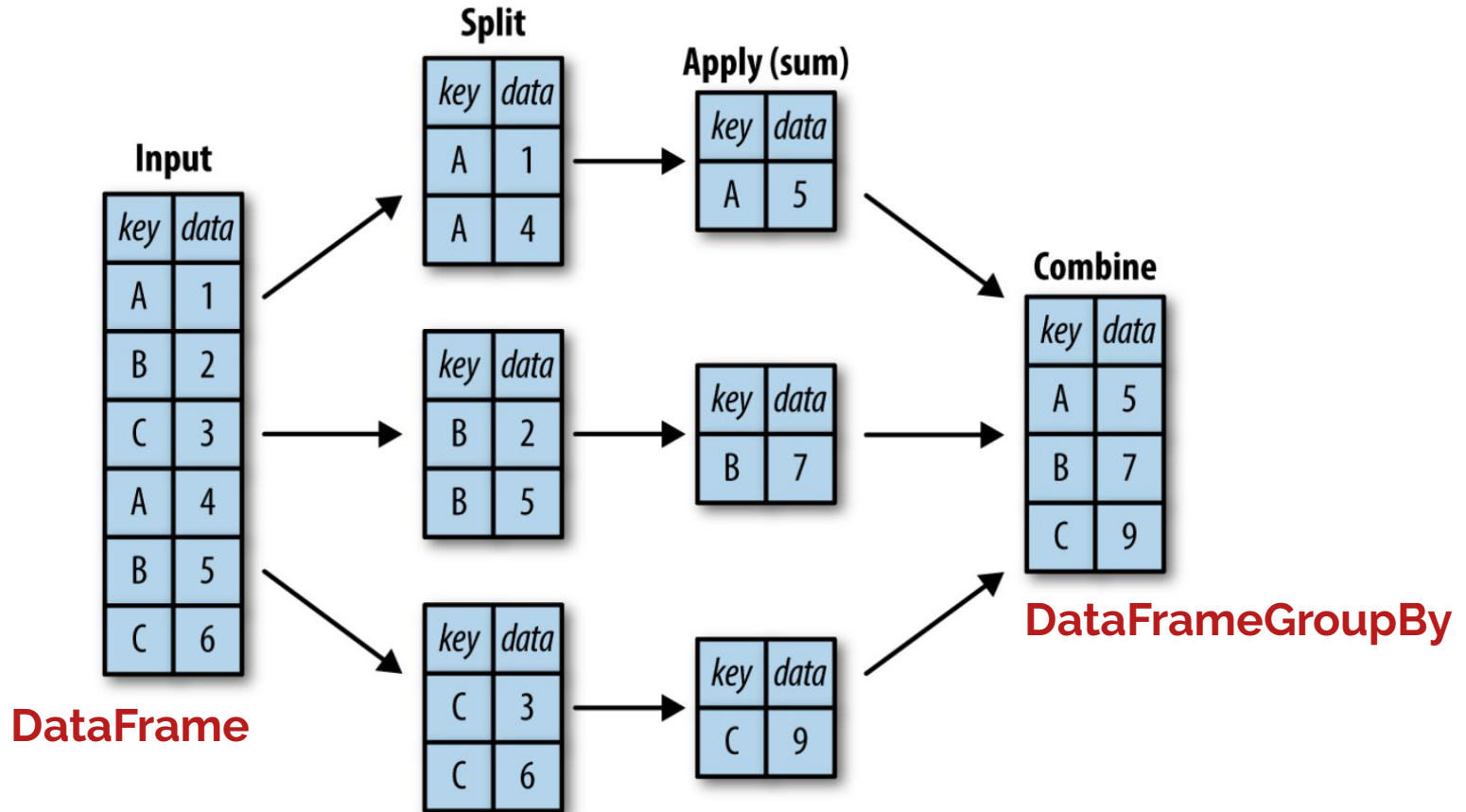
- groupby
- pivot_tables

Método que encapsula una función secuencial:

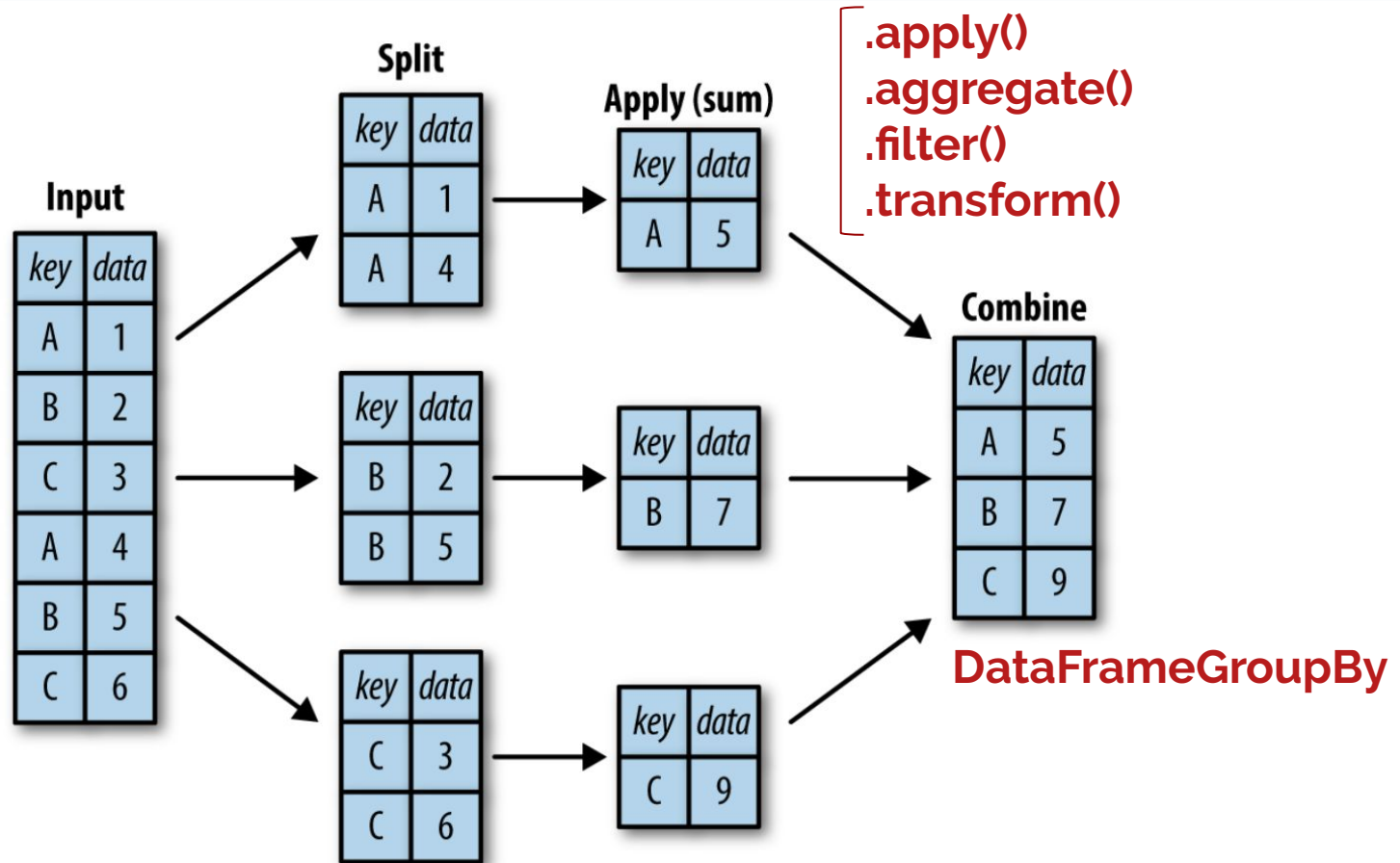
***split, apply, combine* => dividir + aplicar + combinar**

- ***split*** implica dividir y reagrupar un DataFrame en base a una determinada key.
- ***apply*** supone computar alguna función (generalmente, alguna agregación, transformación o filtro) sobre los grupos constituidos en el paso anterior.
- ***combine*** hace un "merge" de los resultados de dichas operaciones en un nuevo array.

df.groupby('key')



df.groupby('key').apply(sum)



Practica Guiada 1



- Permite **ordenar, contar, sumar o aplicar cualquier otra función** de sumariazación sobre el **DataFrame** original, retornando un nuevo DataFrame.
- Los niveles de agregación se guardan con indexación jerárquica: **df.MultiIndex**
- Se puede manipular la visualización a partir de métodos como **.stack()** y **.unstack()**

df.pivot_table(values = [columna(s) cuyos valores quiero resumir],
index = [columna(s) que sirve(n) para agrupar por filas],
columns = [columna(s) que sirve(n) para agrupar por columnas],
aggfunc = [funciones a calcular, listas o diccionarios (media),
fill_value=None, **dropna**=True, **margins**=False, **margins_name**='All'])

	foo	bar	baz	zoo
0	one	A	1	x
1	one	B	2	y
2	one	C	3	z
3	two	A	4	q
4	two	B	5	w
5	two	C	6	t



bar	A	B	C
foo			
one	1	2	3
two	4	5	6

df2

		A	B
first	second		
bar	one	1	2
	two	3	4
baz	one	5	6
	two	7	8

MultIndex

stacked = df2.stack()

first	second		
bar	one	A	1
		B	2
	two	A	3
		B	4
baz	one	A	5
		B	6
	two	A	7
		B	8

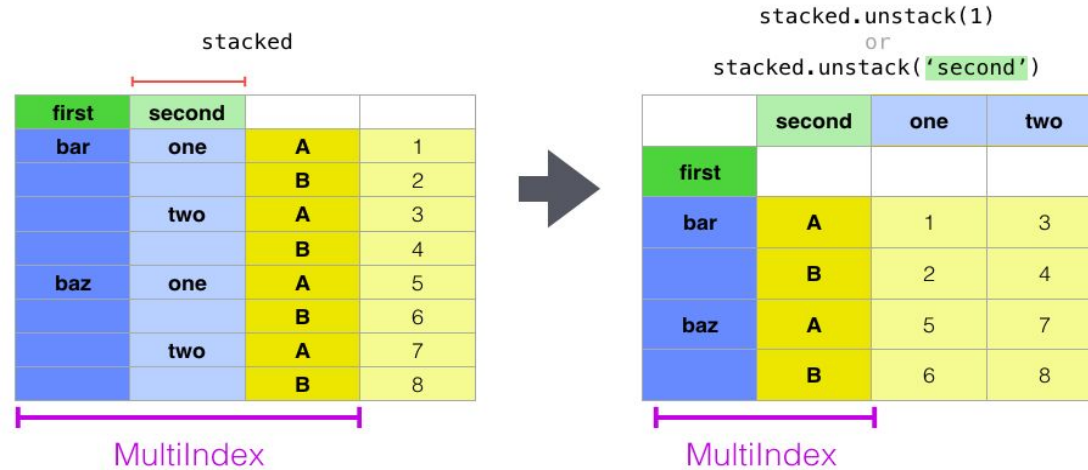
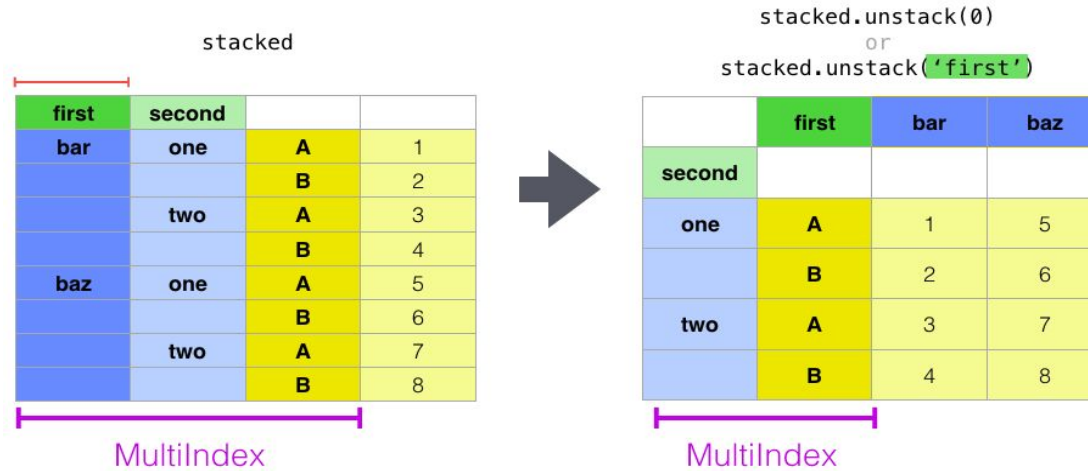
MultIndex

stacked.unstack()

		A	B
first	second		
bar	one	1	2
	two	3	4
baz	one	5	6
	two	7	8

MultIndex

.pivot_table()



df3

	first	last	height	weight
0	John	Doe	5.5	130
1	Mary	Bo	6.0	150



df3.melt(id_vars=['first', 'last'])

	first	last	variable	value
0	John	Doe	height	5.5
1	Mary	Bo	height	6.0
2	John	Doe	weight	130
3	Mary	Bo	weight	150

Practica Guiada 2

