

Trayecto Datos

Clase 8: Agrupaciones y uniones



Agrupaciones

Es muy útil en algunos casos realizar agrupaciones sobre los DataFrame que sirve para ver los datos y organizarlos de otro manera.

Para eso utilizaremos un método de pandas llamado *groupby* que permite agrupar datos en función de los demás



Groupby

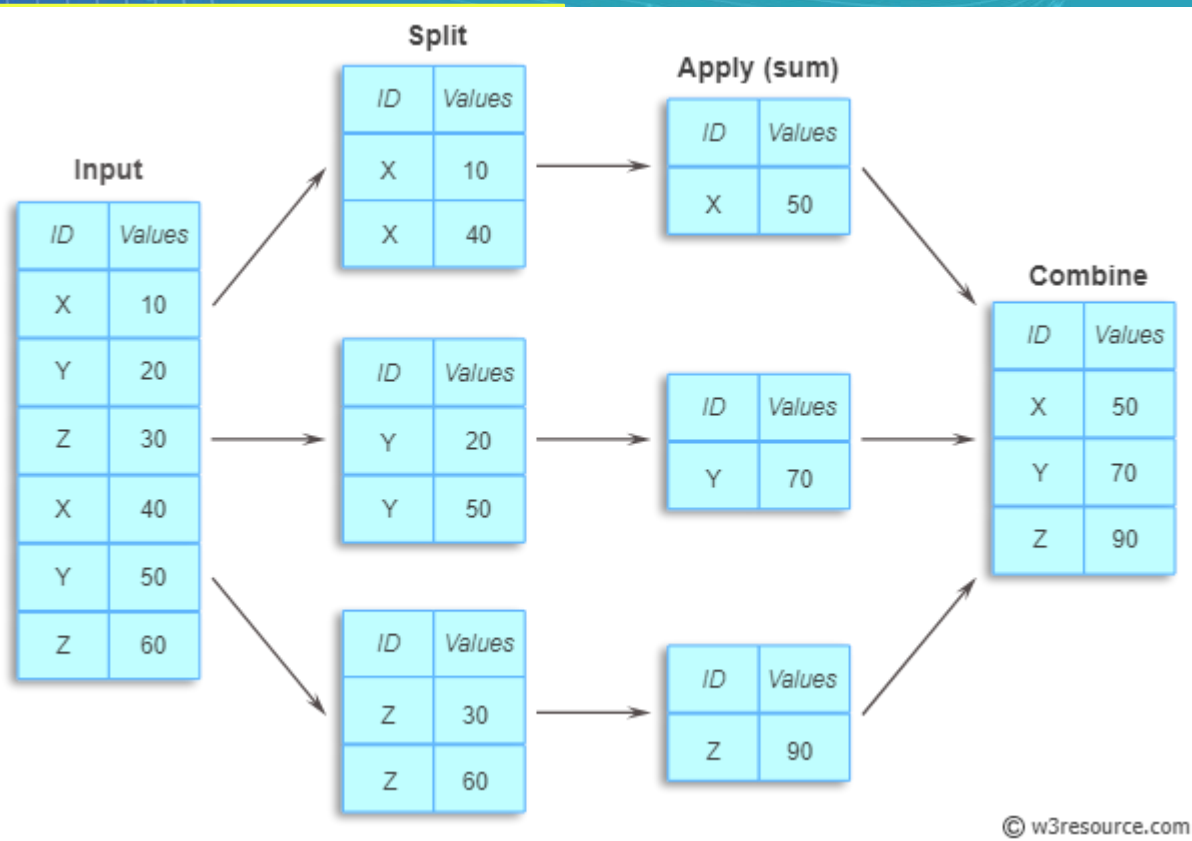
Podemos describir las operaciones sobre grupos con el término *split-apply-combine*

En la primera etapa del proceso, los datos en un objeto pandas (una instancia de Series o de DataFrame) se dividen en grupos (*split*) en base a una o más keys que definimos. Esta división se lleva a cabo por filas ($\text{axis} = 0$) o por columnas ($\text{axis} = 1$).

Como segunda etapa, aplicamos una función a cada uno de los grupos (*apply*) dando como resultado un nuevo valor por grupo.

Como último paso, los resultados de la aplicación de la función en cada uno de los grupos se combina en un objeto resultado (*combine*).

Groupby



Groupby

Para realizar la agrupación en primer lugar se define por que columna se va a realizar la agrupación

Columna1	Columna2
A	10
A	20
B	50
B	10
C	60
C	20



```
DataFrame.groupby("Columna1")
```



Columna1
A
B
C

Groupby

Luego se realiza una operación sobre los datos de una columna por ejemplo suma (*sum()*), promedio (*mean()*), contar (*count()*), etc.

Columna1
A
B
C



```
DataFrame.groupby("Columna1")["Columna2"].sum()
```



Columna1	Columna2
A	30
B	60
C	80

Uniones – *join* / *merge* / *concat*

Estas operaciones se utilizan para unir dos DataFrame o Series.

Concat - Axis 0

df1		A	B	C	D	df2		A	B	C	D
	0	A0	B0	C0	D0		2	A2	B2	C2	D2
	1	A1	B1	C1	D1		3	A3	B3	C3	D3
Concat											
		A	B	C	D						
0		A0	B0	C0	D0						
1		A1	B1	C1	D1						
2		A2	B2	C2	D2						
3		A3	B3	C3	D3						

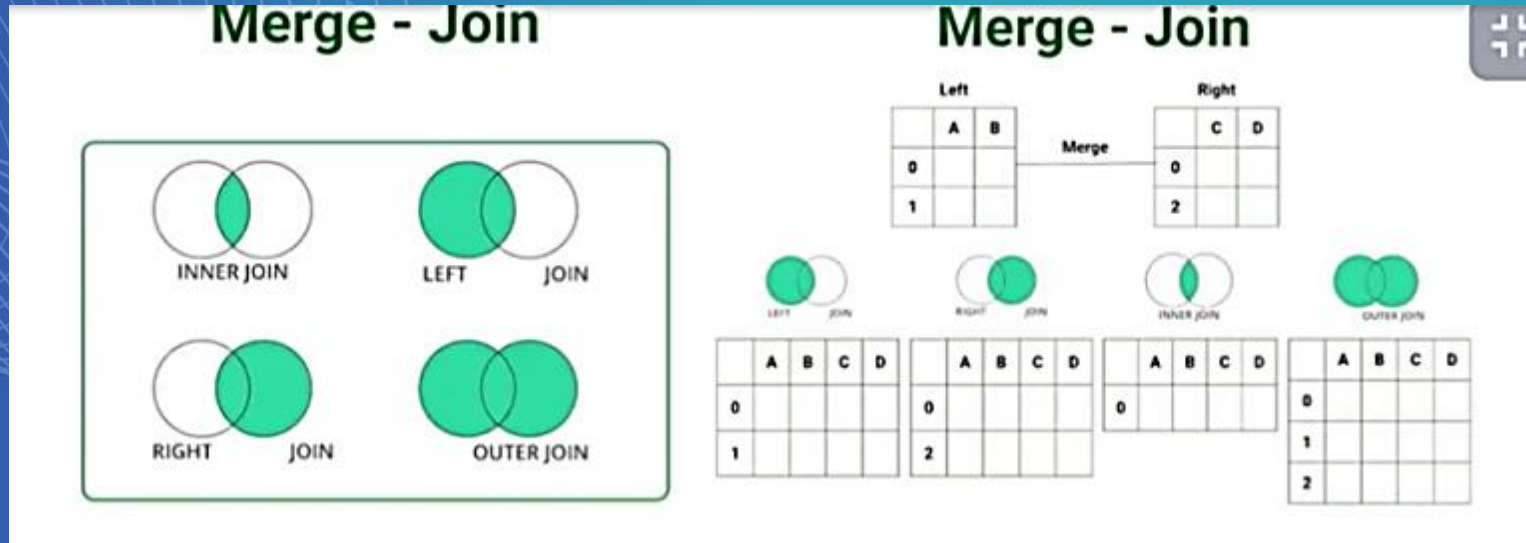
Concat - Axis 1

df1		A	B	C	D	df2		C	D	E	F
	0	A0	B0	C0	D0		1	C1	D1	E1	F1
	1	A1	B1	C1	D1		2	C2	D2	E2	F2
Concat											
		A	B	C	D	C	D	E	F		
0		A0	B0	C0	D0	NaN	NaN	NaN	NaN		
1		A1	B1	C1	D1	C1	D1	E1	F1		
2		NaN	NaN	NaN	NaN	C2	D2	E2	F2		

Fuente: <https://www.linkedin.com/pulse/entiende-los-join-con-python-luis-felipe-castro-calder%C3%B3n/?originalSubdomain=es>

Uniones – *join* / *merge* / *concat*

Estas operaciones se utilizan para unir dos DataFrame o Series.



Fuente: <https://www.linkedin.com/pulse/entiende-los-join-con-python-luis-felipe-castro-calder%C3%B3n/?originalSubdomain=es>

¡MUCHAS GRACIAS!

