

DigitalHouse >
Coding School

DATA SCIENCE

UNIDAD 1
MÓDULO 2

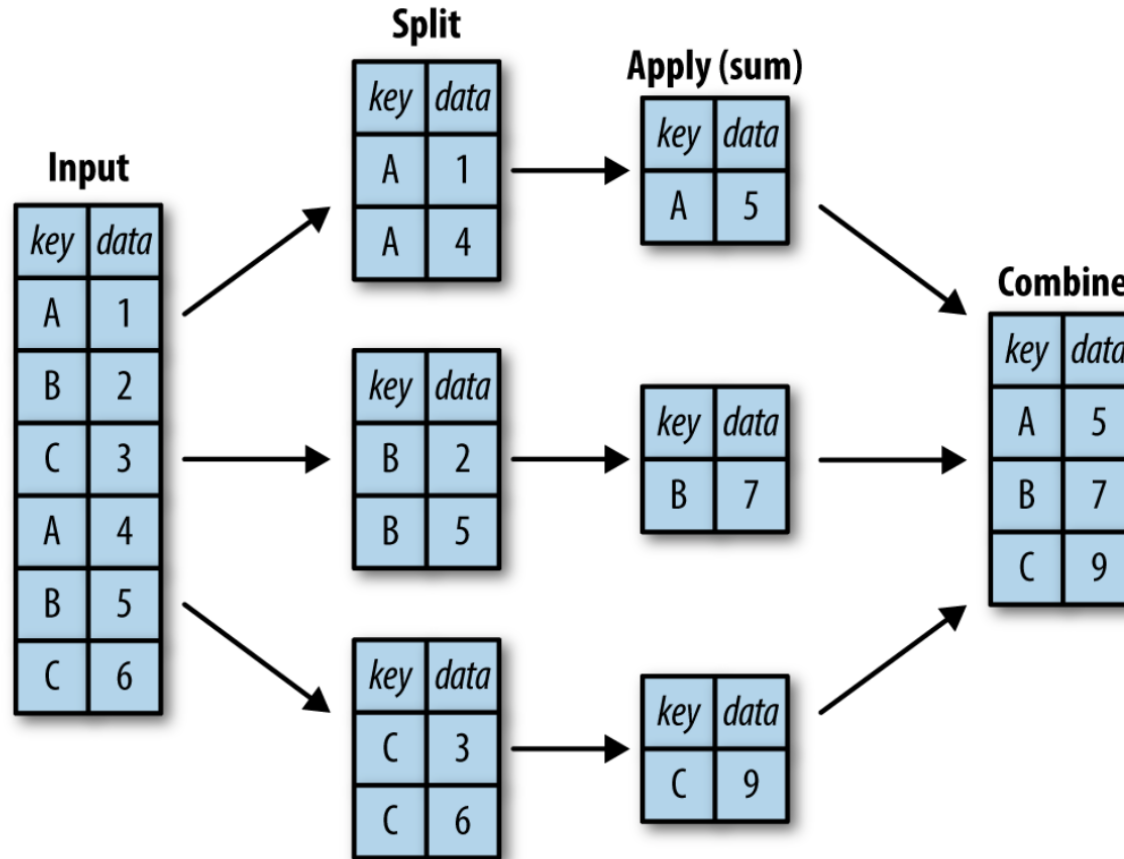
Introducción a Pandas 2

Abril 2017

Pandas



GroupBy y sumarización



Práctica Guiada 1

- Una **tabla pivot** es una herramienta utilizada en sistemas de visualización de datos para sumarizar los mismos
- Permite **ordenar, contar, sumar o aplicar otra función** de sumarización de los de una tabla inicial y genera una nueva con el resultado
- Hay muchas formas de visualizar el resultado pero mayoritariamente los valores de las filas de una columna pasan a ser columnas por eso se denomina **pivot**
- Los DataFrame de pandas poseen el método **pivot_table()** que genera otro DataFrame de salida con el resultado del pivot según los argumentos pasados a la función.

Datos en crudo

Fecha	Vendedor	Unidades	Monto
10/03/2017	Juan	10	500
10/03/2017	Pedro	4	200
11/03/2017	Juan	5	500
11/03/2017	Pedro	2	50

PIVOT 1

Vendedor	Cantidad Total
Juan	15
Pedro	6

PIVOT 2

Métrica	Juan	Pedro
Monto Total	1000	250

Ejemplo de uso de pivot_table

<u>ix</u>	<u>Item</u>	<u>CType</u>	<u>USD</u>	<u>EU</u>
<u>0</u>	Item0	Gold	1	1
<u>1</u>	Item0	Bronze	2	2
<u>2</u>	Item0	Gold	3	3
<u>3</u>	Item1	Silver	4	4

<u>ix=Item</u>	<u>Bronze</u>	<u>Gold</u>	<u>Silver</u>
<u>Item0</u>	2	2 = mean(1,3)	NaN
<u>Item1</u>	NaN	NaN	4

```
d.pivot_table(index='Item', columns='CType', values='USD', aggfunc=np.mean)
```

Práctica Guiada 2

LAB