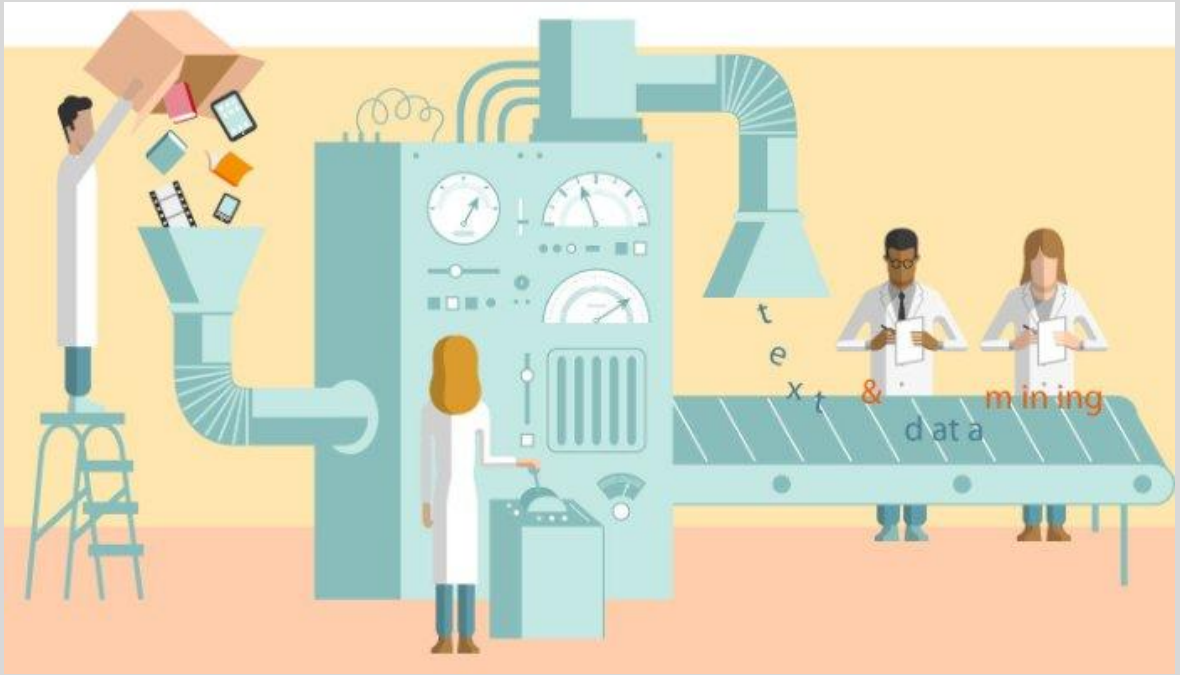


DigitalHouse > Coding School

DATA SCIENCE

UNIDAD 1
MÓDULO 2

Joins con Pandas

Marzo 2017

Pandas

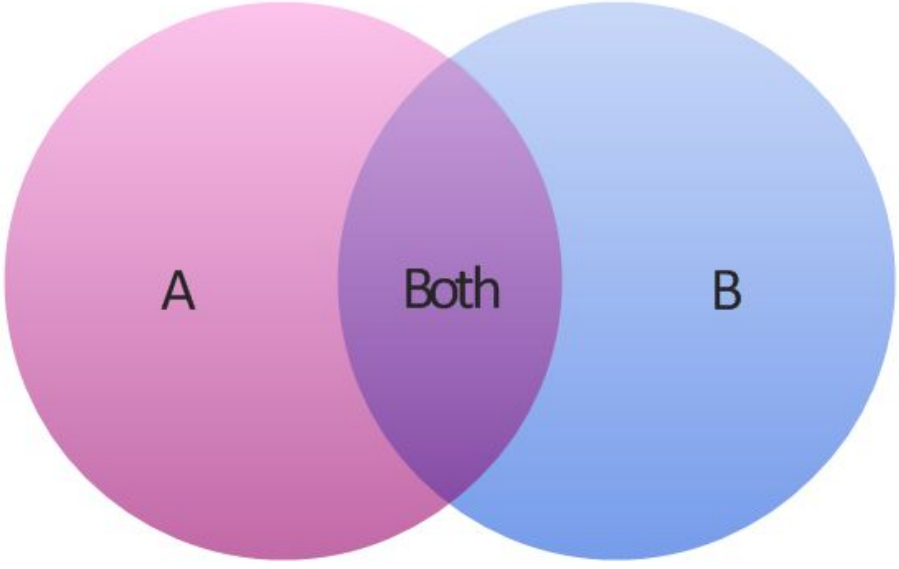
1. Aprender a unir distintas fuentes de información con Pandas

2. Aprender los distintos tipos de Joins, para qué sirven y cómo implementarlos con Pandas

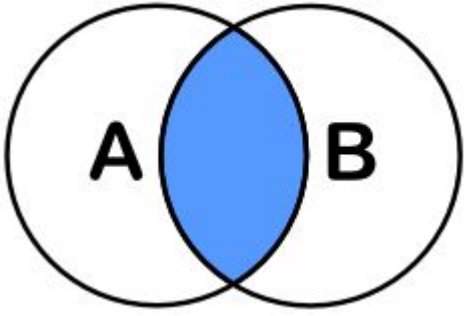
3. Aprender a manejar la función shift() para generar rezagos sobre los datos

Práctica Guiada (Parte I)

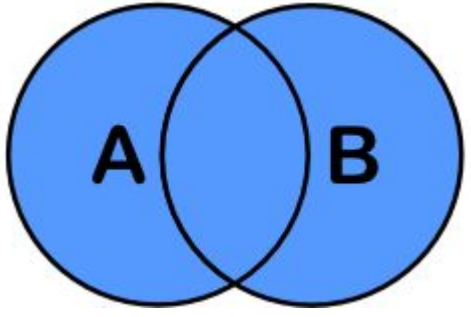
Conjuntos y Diagramas de Venn



JOINS con Diagramas de Venn

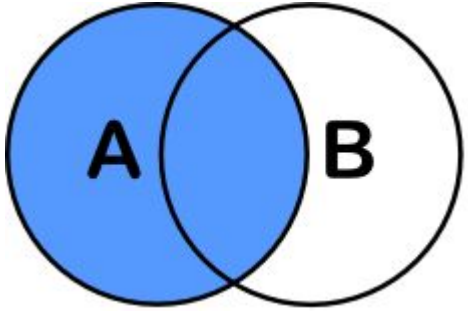


INNER JOIN

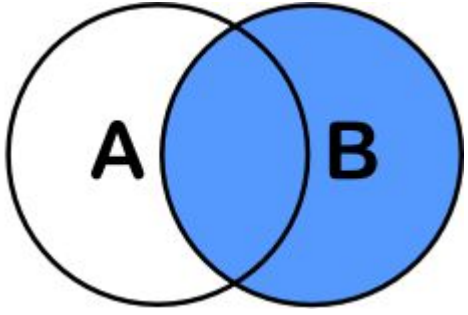


FULL OUTER JOIN

JOINS con Diagramas de Venn



LEFT JOIN



RIGHT JOIN

INNER JOIN con DataFrame

left

	A	B	key
0	A0	B0	K0
1	A1	B1	K1
2	A2	B2	K2
3	A3	B3	K3

right

	C	D	key
0	C0	D0	K0
1	C1	D1	K1
2	C2	D2	K2
3	C3	D3	K3

Result

	A	B	key	C	D
0	A0	B0	K0	C0	D0
1	A1	B1	K1	C1	D1
2	A2	B2	K2	C2	D2
3	A3	B3	K3	C3	D3

PANDAS BENCHMARKS

DigitalHouse > Coding School

pandas vs R merge benchmarks

Many-to-one

	pandas	data.table	plyr	base::merge
inner	1	5.905	6.35	13.29
outer	1	10.05	9.209	20.25
left	1	2.849	5.918	14.93
right	1	2.05	2.923	16.91

Many-to-many

	pandas	data.table	plyr	base::merge
inner	1	5.194	5.223	18.87
outer	1	10	6.903	33.75
left	1	2.528	4.688	24.46
right	1	1.681	2.05	25.24

Pandas vs sqlite3

	sqlite3	pandas
inner	0.02328	0.01799
outer	0.02324	0.01943
left	0.02324	0.01923

Práctica Guiada (Parte II)

Práctica Independiente

Laboratorio