Analytics Jumpstart

Joining Dataframes

Nashville Software School

For today

- More pandas
  - Merging vs. Concatenating
  - Aggregating
  - groupby

Merging two DataFrames

pd.merge(<df1>, <df2>, on = <col or list of cols to join on>, how = <join_type>*)*

left

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

right

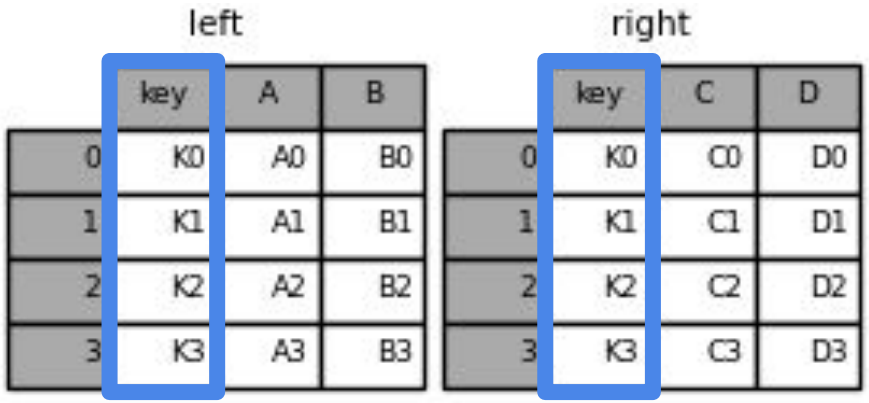
	key	C	D
0	K0	C0	D0
1	K1	C1	D1
2	K2	C2	D2
3	K3	C3	D3

- Need one or more “key” columns to join on
- Pastes matching rows together along the key column(s)

NASHVILLE SOFTWARE SCHOOL

Merging two DataFrames

pd.merge(<df1>, <df2>, on = <col or list of cols to join on>, how = <join_type>*)*



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

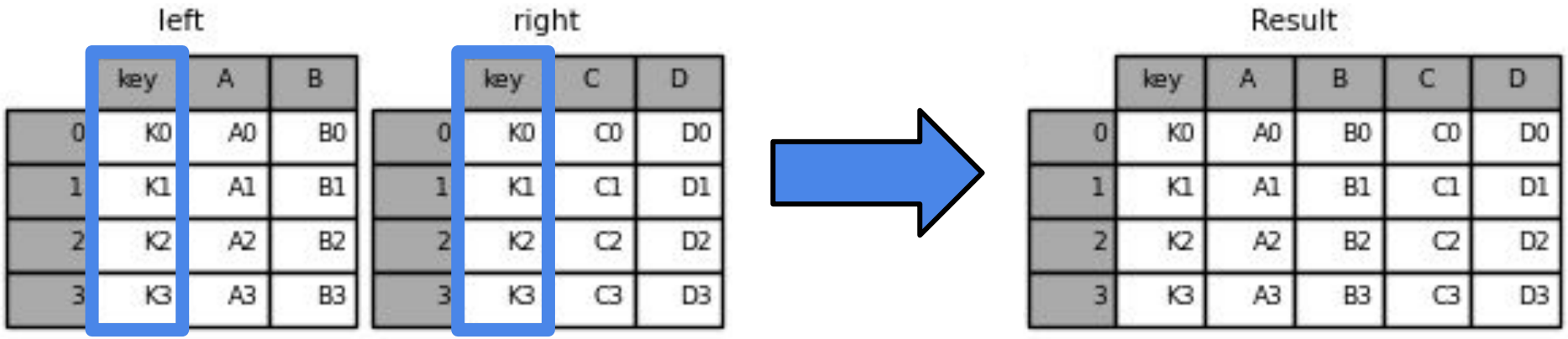
	key	C	D
0	K0	C0	D0
1	K1	C1	D1
2	K2	C2	D2
3	K3	C3	D3

- Need one or more “key” columns to join on
- Pastes matching rows together along the key column(s)

NASHVILLE SOFTWARE SCHOOL

Merging two DataFrames

pd.merge(*<df1>, <df2>, on = <col or list of cols to join on>, how = <join_type>*)



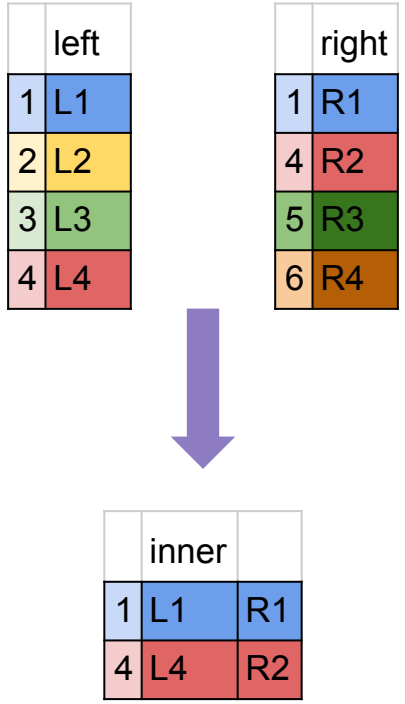
- **Need one or more “key” columns to join on**
- **Pastes matching rows together along the key column(s)**

NASHVILLE SOFTWARE SCHOOL

INNER JOIN

An **INNER JOIN** keeps **only the rows that have matching values in both tables**.

This is the default type of join when using *pd.merge()*.

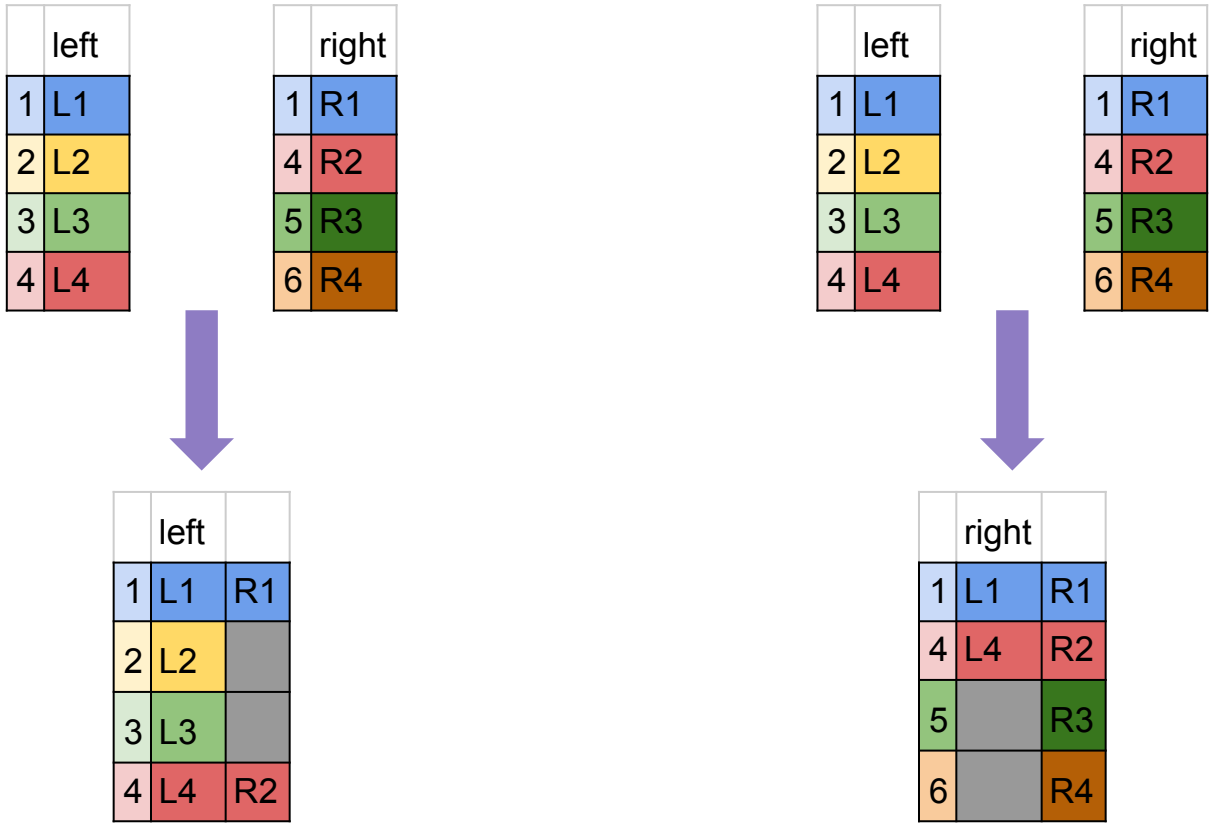


NASHVILLE SOFTWARE SCHOOL

LEFT JOIN and RIGHT JOIN

A **LEFT JOIN** keeps all rows from the left table and all matching rows from the right table. A **RIGHT JOIN** works similarly, except all rows from the right table are kept.

how = "left" or *how = "right"*

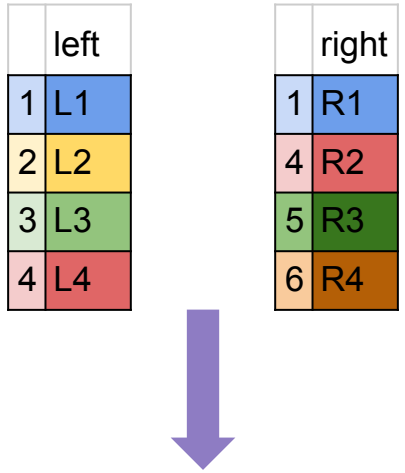


NASHVILLE SOFTWARE SCHOOL

OUTER/FULL JOIN

AN **OUTER JOIN** keeps all rows from both tables.

how = “outer”



NASHVILLE SOFTWARE SCHOOL

Concatenating DataFrames

pd.concat([<df1>, <df2>, <df3>]*)*

- Takes two or more DataFrames that have the same columns.
- Combines them by stacking vertically (can also be done horizontally).

df1

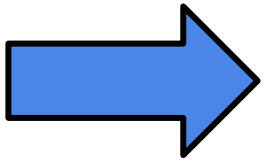
	A	B	C	D
0	A0	B0	C0	D0
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	A3	B3	C3	D3

df2

	A	B	C	D
4	A4	B4	C4	D4
5	A5	B5	C5	D5
6	A6	B6	C6	D6
7	A7	B7	C7	D7

df3

	A	B	C	D
8	A8	B8	C8	D8
9	A9	B9	C9	D9
10	A10	B10	C10	D10
11	A11	B11	C11	D11



Result

	A	B	C	D
0	A0	B0	C0	D0
1	A1	B1	C1	D1
2	A2	B2	C2	D2
3	A3	B3	C3	D3
4	A4	B4	C4	D4
5	A5	B5	C5	D5
6	A6	B6	C6	D6
7	A7	B7	C7	D7
8	A8	B8	C8	D8
9	A9	B9	C9	D9
10	A10	B10	C10	D10
11	A11	B11	C11	D11

NASHVILLE SOFTWARE SCHOOL

Questions?