

Intro to Pandas



SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority



Agenda



What is Pandas



Pandas Core Features



Pandas Data Structure



Pandas Data Types



Pandas Files I/O



Pandas Joins/Merge



Grouping and Aggregations



What is Pandas

Pandas is a powerful Python library for data manipulation and analysis by reading most data sources and representing it in tabular format. It supports SQL-like merging, joining, and enables time series analysis with built-in functions for date manipulation and statistical operations. It facilitates data grouping, aggregation, and basic data visualization, easily integrating with libraries for more advanced graphics.



Pandas

Core Features

Data Structures:

- Core structures include DataFrames for tabular data and Series for single-column data of any type.

File I/O:

- Reads and writes various file formats (CSV, Excel, SQL, JSON), facilitating data exchange.

Merging and Joining:

- Supports SQL-like merging, joining, and concatenating from different sources.

Data Manipulation:

- Offers functions for cleaning, filtering, and transforming data, ideal for large datasets.

Data Grouping and Aggregation:

- Enables sophisticated operations like group-wise computations

Data Visualization:

- Supports basic plotting and integrates with libraries like Matplotlib and Seaborn for advanced visuals.



▶▶ Pandas Data Structure

Series

	Prices
1	126
2	240
3	320
4	422
5	519
6	616

DataFrame

	Prices	Car	Car Type
1	126	Hyundai	Sedan
2	240	Kia	Crossover
3	320	Nissan	Crossover
4	422	Opel	Sedan
5	519	Jeep	Crossover
6	616	Porsche	Sport

Index

Columns

Header

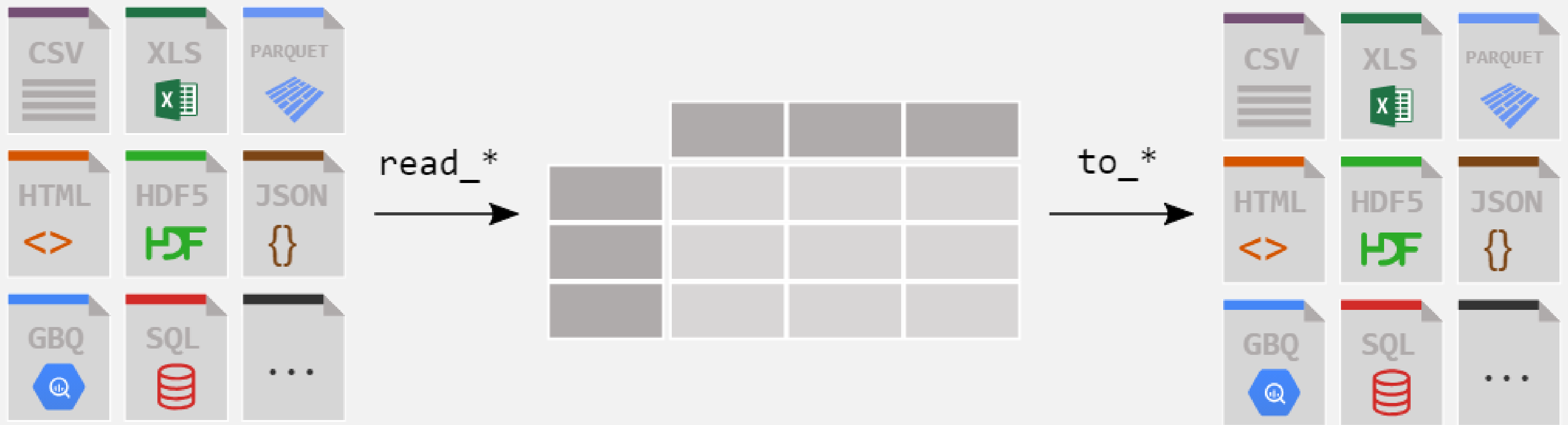
Rows



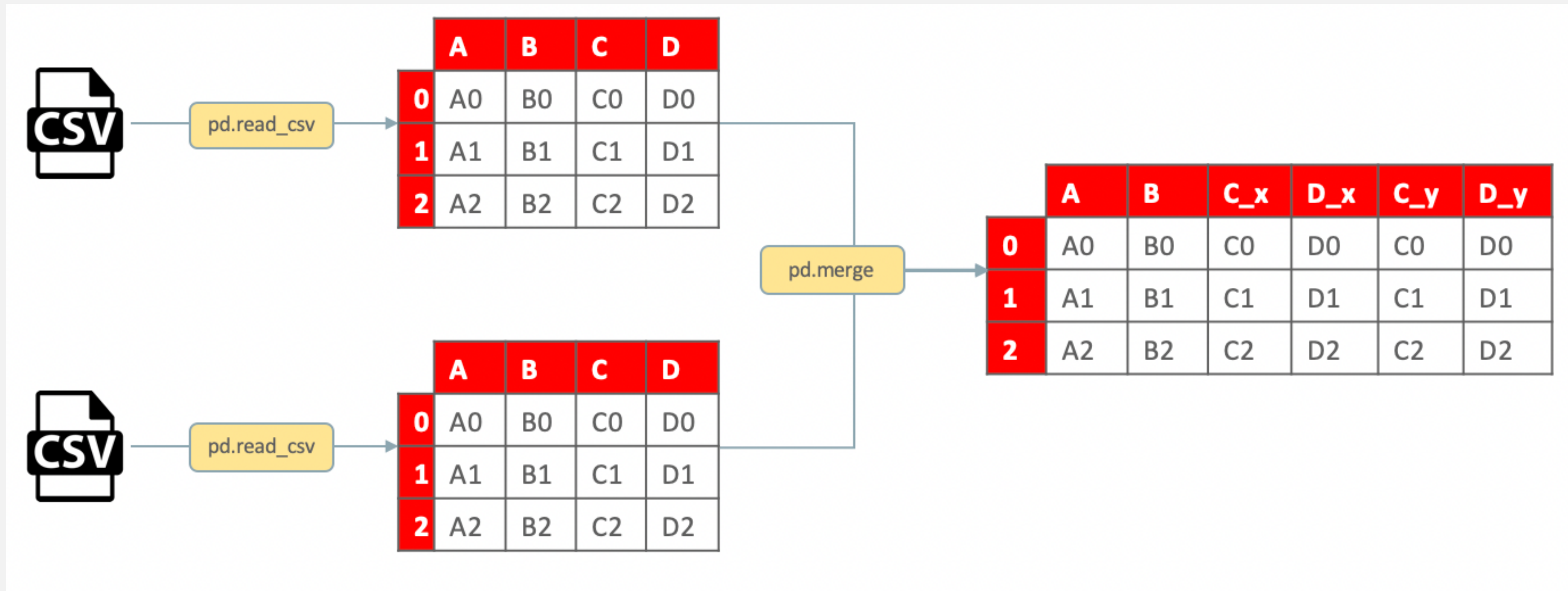
▶▶ Pandas Data Types

Pandas dtype	Python type	NumPy type	Usage
object	str or mixed	string_, unicode_, mixed types	Text or mixed numeric and non-numeric values
int64	int	int_, int8, int16, int32, int64, uint8, uint16, uint32, uint64	Integer numbers
float64	float	float_, float16, float32, float64	Floating point numbers
bool	bool	bool_	True/False values
datetime64	NA	datetime64[ns]	Date and time values
timedelta[ns]	NA	NA	Differences between two datetimes
category	NA	NA	Finite list of text values

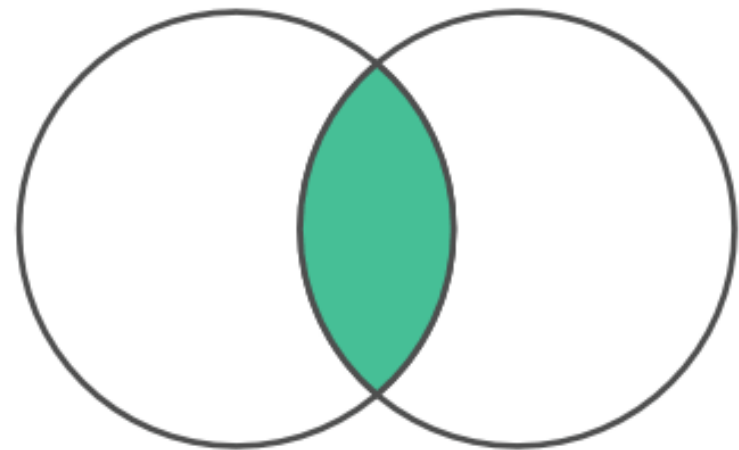
▶▶ Pandas Files I/O



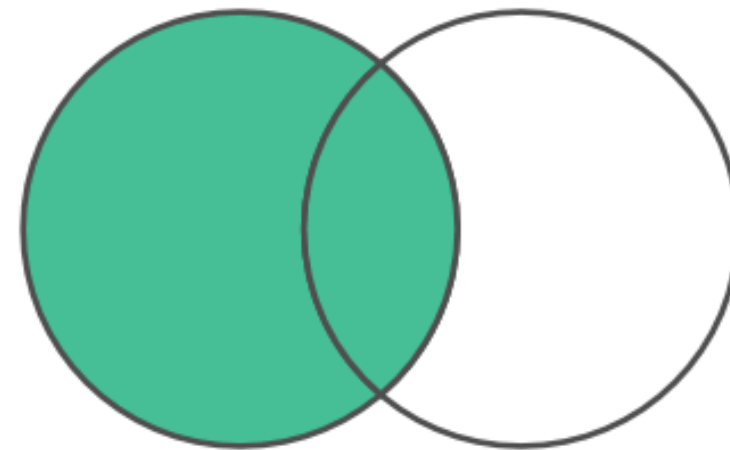
Pandas Merging



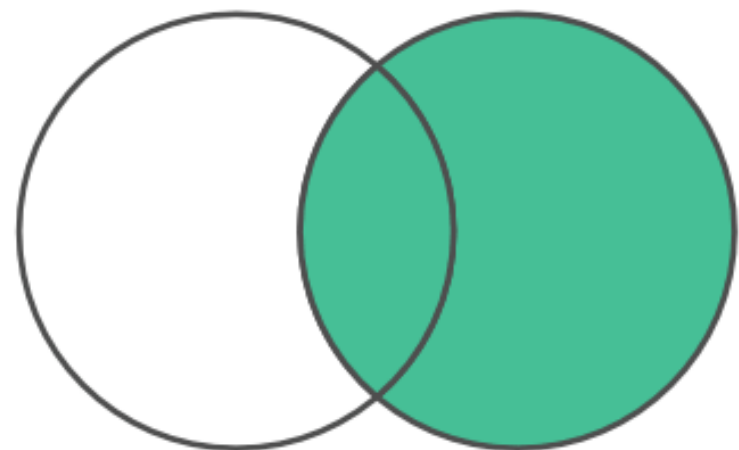
▶▶ Pandas Merging



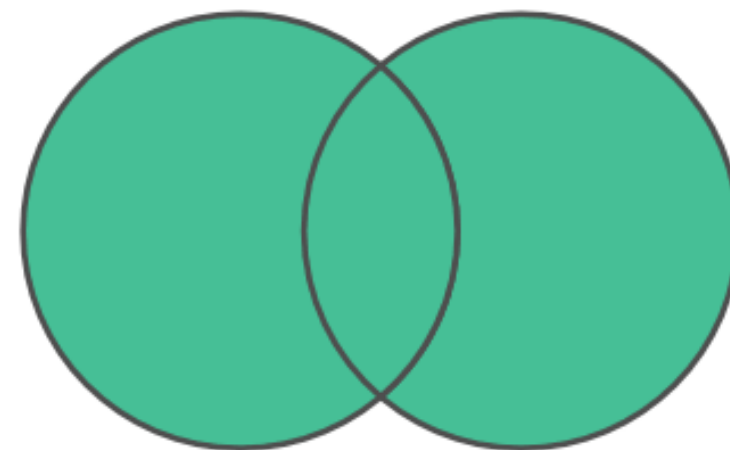
INNER JOIN



LEFT OUTER JOIN



RIGHT OUTER JOIN



FULL OUTER JOIN

▶▶ Data Grouping and Aggregation

	species	sepal_length	sepal_width	petal_length	petal_width
0	setosa	5.1	3.5	1.4	0.2
1	setosa	4.9	3.0	1.4	0.2
2	setosa	4.7	3.2	1.3	0.2
3	setosa	4.6	3.1	1.5	0.2
4	setosa	5.0	3.6	1.4	0.2
50	versicolor	7.0	3.2	4.7	1.4
51	versicolor	6.4	3.2	4.5	1.5
52	versicolor	6.9	3.1	4.9	1.5
53	versicolor	5.5	2.3	4.0	1.3
54	versicolor	6.5	2.8	4.6	1.5
100	virginica	6.3	3.3	6.0	2.5
101	virginica	5.8	2.7	5.1	1.9
102	virginica	7.1	3.0	5.9	2.1
103	virginica	6.3	2.9	5.6	1.8
104	virginica	6.5	3.0	5.8	2.2

SUM

SUM

SUM

	species	sepal_length	sepal_width	petal_length	petal_width
	setosa	24.3	16.4	7.0	1.0
	versicolor	32.3	14.6	22.7	7.2
	virginica	32.0	14.9	28.4	10.5



شكراً لكم



SDAIA
الهيئة السعودية للبيانات
والذكاء الاصطناعي
Saudi Data & AI Authority