The Pandas Bootcamp

master your data.

Tabular Data (Pandas DataFrame)

Column

Player	Nationality	Club	World_Champion	Height	Goals_2018
Lionel Messi	Argentina	FC Barcelona	False	1.70	45
Cristiano Ronaldo	Portugal	Juventus FC	False	1.87	44
Neymar Junior	Brasil	Paris SG	False	1.75	28
Kylian Mbappe	France	Paris SG	True	1.78	21
Manuel Neuer	Germany	FC Bayern	True	1.93	0

Element / Value / Entry

Row

Subset / Slice

Tabular Data (Pandas DataFrame)

2 Axis

(Row) Index Name

Column Label / Column Header

Columns / Column Index

(Row) Index

Index Label

Player	Nationality	Club	World_Champion	Height	Goals_2018
Lionel Messi	Argentina	FC Barcelona	False	1.70	45
Cristiano Ronaldo	Portugal	Juventus FC	False	1.87	44
Neymar Junior	Brasil	Paris SG	False	1.75	28
Kylian Mbappe	France	Paris SG	True	1.78	21
Manuel Neuer	Germany	FC Bayern	True	1.93	0

Optimal Data Organization in DataFrames

Characteristics / Variables / Features
(e.g. height) →

Observations
(e.g. football players)

Player	Nationality	Club	World_Champion	Height	Goals_2018
Lionel Messi	Argentina	FC Barcelona	False	1.70	45
Cristiano Ronaldo	Portugal	Juventus FC	False	1.87	44
Neymar Junior	Brasil	Paris SG	False	1.75	28
Kylian Mbappe	France	Paris SG	True	1.78	21
Manuel Neuer	Germany	FC Bayern	True	1.93	0
	string	string	bool	float	integer

optimal: one datatype per column!