

# Pandas

Getting started

# What is pandas?

- Name is derived from “panel data”
- Open source, originally written by Wes Mackinney
- Software of choice across several domains (neuroscience, finance, economics, statistics, web analytics, ...)
- Provides additional functionality of high-level spreadsheets and databases
- Built on top of NumPy

# DataFrame object as Main goal

DATA (table format)

- .csv .xls .sql .json



pandas



DataFrame object

- Easier to work with than lists, dictionaires

# Components of a DataFrame

The diagram illustrates the components of a DataFrame using a table of mountain data. A blue bracket on the left side of the table is labeled "index labels", pointing to the row indices 0 through 9. A red bracket at the top of the table is labeled "column names", pointing to the header row. An orange bracket on the right side of the table is labeled "data", pointing to the data rows. The table itself has 9 columns: Mountain, Height (m), Range, Coordinates, Parent mountain, First ascent, Ascents bef. 2004, and Failed attempts bef. 2004. The data rows are as follows:

	Mountain	Height (m)	Range	Coordinates	Parent mountain	First ascent	Ascents bef. 2004	Failed attempts bef. 2004
0	Mount Everest / Sagarmatha / Chomolungma	8848	Mahalangur Himalaya	27°59'17"N 86°55'31"E	NaN	1953	>>145	121.0
1	K2 / Qogir / Godwin Auster	8611	Baltoro Karakoram	35°52'53"N 76°30'48"E	Mount Everest	1954	45	44.0
2	Kangchenjunga	8586	Kangchenjunga Himalaya	27°42'12"N 88°08'51"E	Mount Everest	1955	38	24.0
3	Lhotse	8516	Mahalangur Himalaya	27°57'42"N 86°55'59"E	Mount Everest	1956	26	26.0
4	Makalu	8485	Mahalangur Himalaya	27°53'23"N 87°05'20"E	Mount Everest	1955	45	52.0
5	Cho Oyu	8188	Mahalangur Himalaya	28°05'39"N 86°39'38"E	Mount Everest	1954	79	28.0
6	Dhaulagiri I	8167	Dhaulagiri Himalaya	28°41'48"N 83°29'35"E	K2	1960	81	39.0
7	Manaslu	8163	Manaslu Himalaya	28°33'00"N 84°33'35"E	Cho Oyu	1956	49	45.0
8	Nanga Parbat	8126	Nanga Parbat Himalaya	35°14'14"N 74°35'21"E	Dhaulagiri	1953	52	67.0
9	Annapurna I	8091	Annapurna Himalaya	28°35'44"N 83°49'13"E	Cho Oyu	1950	36	47.0

# In addition to DataFrame object

- Tools for reading and writing data
- Data alignment and integrated handling of missing data
- Ability to perform arithmetic operations
- Easy reshaping and pivoting of datasets
- User-friendly operations for merging and joining data
- Ability to handle time series

# Pandas main strenghts

- User-friendly
- Efficient – highly optimized for performance