

# Basic Machine Learning: Data Loading

Oleh: Muhammad Angga Muttaqien | Founder & Mentor

Goal

# Goal

To understand how to load and briefly process data with Pandas DataFrame

1. Load datasets
2. Explore data
3. Describe Data
4. Edit Data

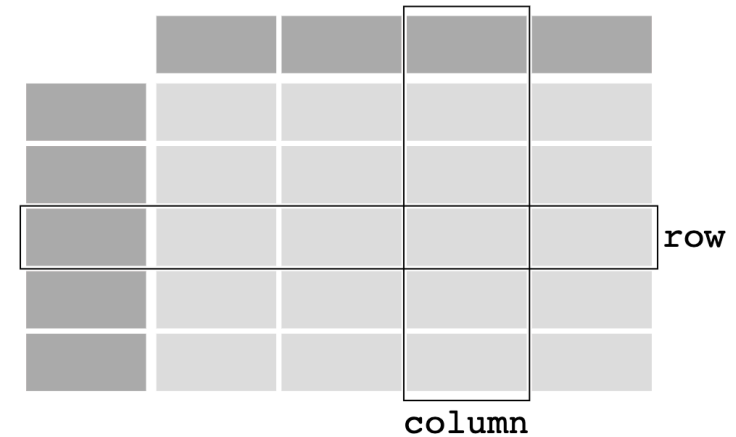


# About Pandas

Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

## pandas data table representation

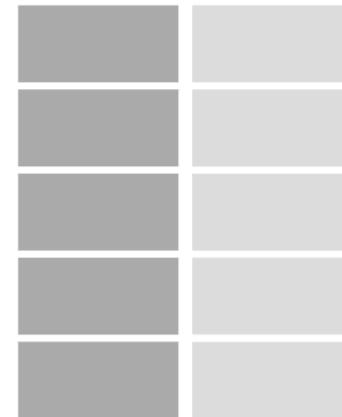
## DataFrame



# About Pandas

Each column in a DataFrame is a Series. When selecting a single column of a pandas DataFrame, the result is a pandas Series.

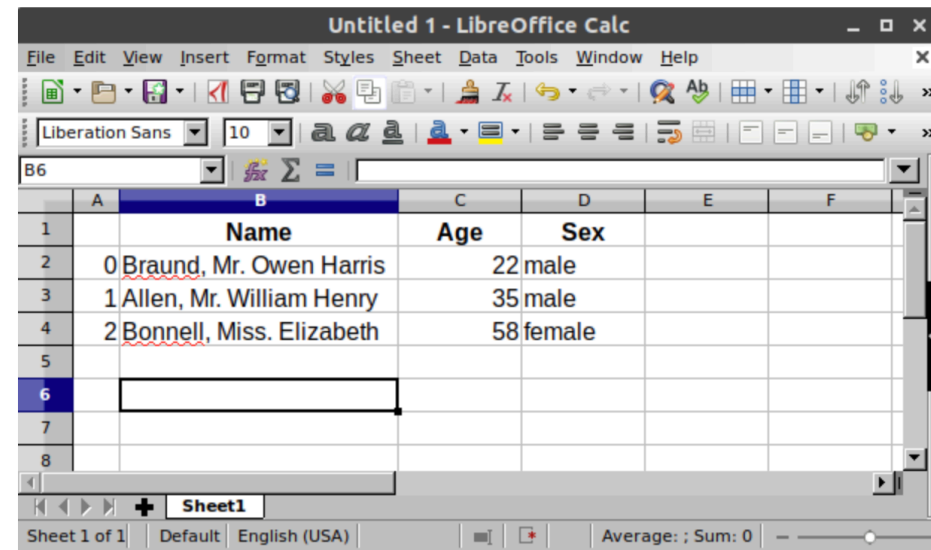
## Series



# About Pandas

A DataFrame is a 2-dimensional data structure that can store data of different types (including characters, integers, floating point values, categorical data and more) in columns. It is similar to a spreadsheet, a SQL table.

In spreadsheet software, the table representation of our data would look very similar:



The screenshot shows the LibreOffice Calc application window titled "Untitled 1 - LibreOffice Calc". The spreadsheet has columns A through F and rows 1 through 8. The data is organized into a table with three columns: "Name" (column B), "Age" (column C), and "Sex" (column D). The rows contain the following data:

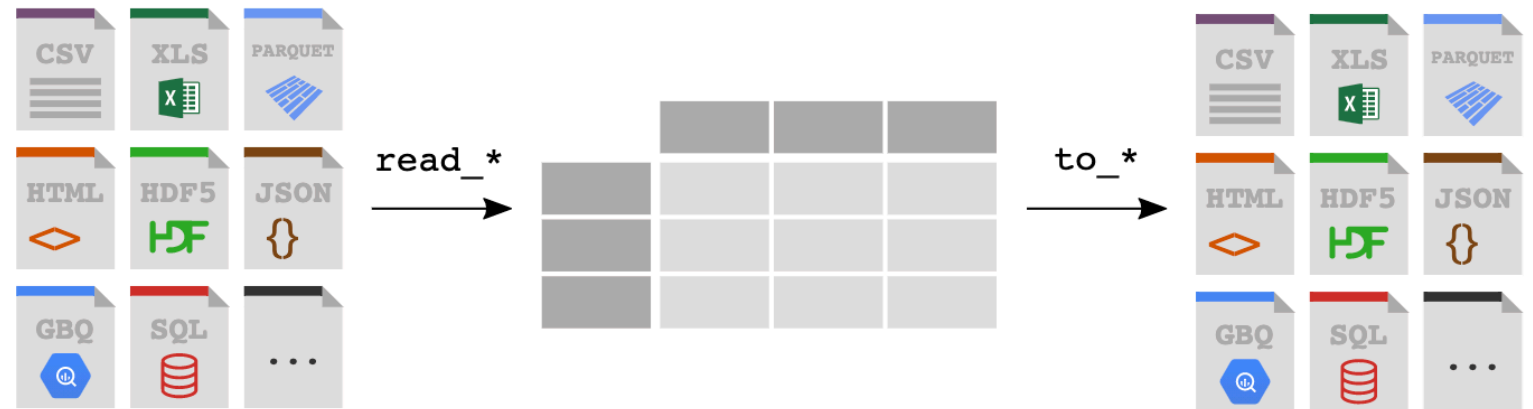
	A	B	C	D	E	F
1		Name	Age	Sex		
2	0	Braund, Mr. Owen Harris	22	male		
3	1	Allen, Mr. William Henry	35	male		
4	2	Bonnell, Miss. Elizabeth	58	female		
5						
6						
7						
8						

The status bar at the bottom indicates "Sheet 1 of 1", "Default", "English (USA)", and "Average: ; Sum: 0".

# About Pandas

By using Pandas, we can work efficiently with the following type of tabular data:

1. CSV file
2. JSON file
3. Txt file
4. Excel file
5. Many others



# About Pandas

By using Pandas, we can read and write the value of the tabular data such as selecting, editing and removing specific value of the specific columns and rows.

How do I select specific columns from a **DataFrame**?

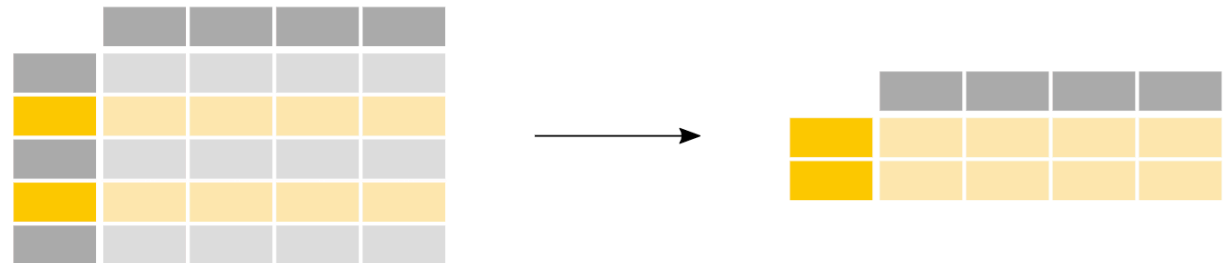




# About Pandas

By using Pandas, we can read and write the value of the tabular data such as selecting, editing and removing specific value of the specific columns and rows.

How do I filter specific rows from a **DataFrame**?



Thanks!