Daedalus Python Workshop

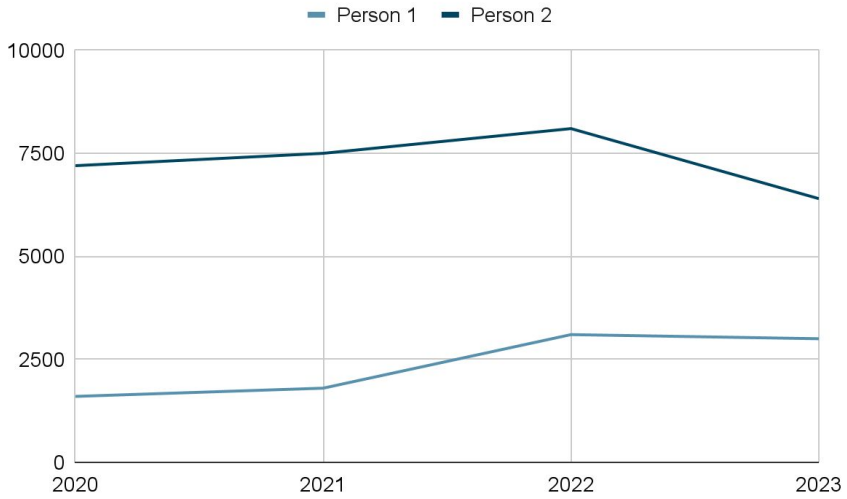
L6: Pandas for Data Analysis

Alexander J. Peterson Santos & Peter Geldner

Garching, 10. December 2025

"Pandas" comes from Panel Data

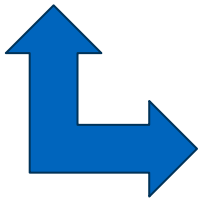
Index	Person	Year	Income	Expenses
0	1	2020	1600	1100
1	1	2021	1800	1100
2	1	2022	3100	1800
3	1	2023	3000	1800
4	2	2020	7200	3000
5	2	2021	7500	3200
6	2	2022	8100	3100
7	2	2023	6400	3300

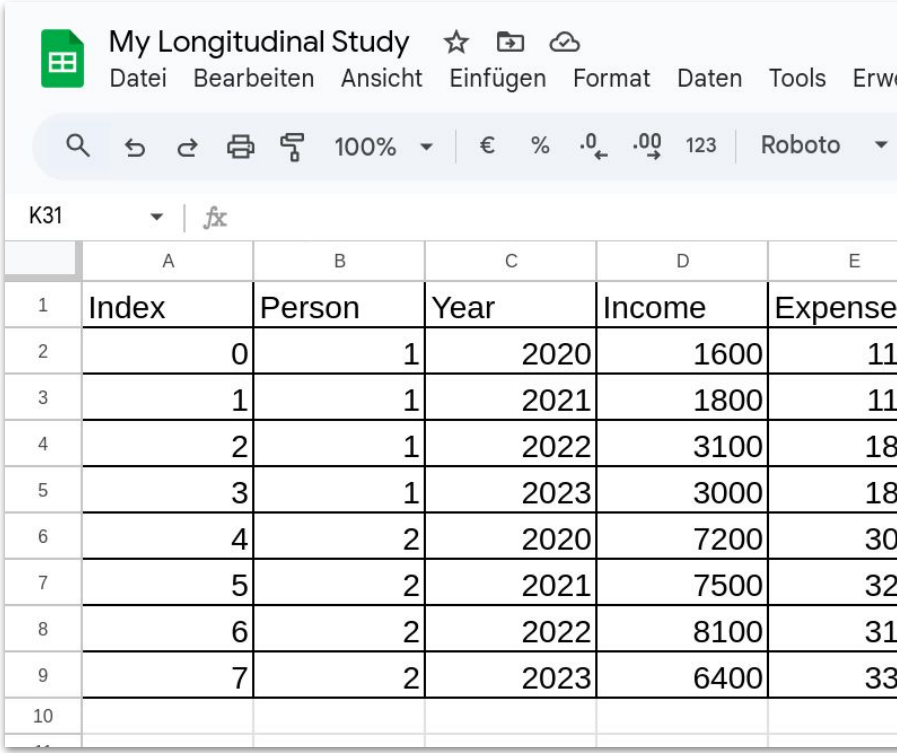


Panel data is **cross-sectional** and **longitudinal**

Pandas is great for Data Science

Index	Person	Year	Income	Expenses
0	1	2020	1600	1100
1	1	2021	1800	1100
2	1	2022	3100	1800
3	1	2023	3000	1800
4	2	2020	7200	3000
5	2	2021	7500	3200
6	2	2022	8100	3100
7	2	2023	6400	3300





	A	B	C	D	E
1	Index	Person	Year	Income	Expense
2	0	1	2020	1600	11
3	1	1	2021	1800	11
4	2	1	2022	3100	18
5	3	1	2023	3000	18
6	4	2	2020	7200	30
7	5	2	2021	7500	32
8	6	2	2022	8100	31
9	7	2	2023	6400	33
10					

Not just panel data

Index	UserID	Registered	Age	Followers
Luke	542542	2013	19	1100
Sara	563732	2016	21	1100
Simon	567236	2009	38	1800
Nico	493763	2021	31	1800
Sabine	659392	2010	46	3000
Moritz	354774	2024	23	3200
Mary	396629	2017	34	3100
Tiago	946385	2019	26	3300

pd.DataFrame

Deals with data in columns and rows

pd.Index

Provides a label for each row
(by default, an enumeration like 0, 1, 2…)

Not just two-dimensional data

Index	Followers
Luke	1100
Sara	1100
Simon	1800
Nico	1800
Sabine	3000
Moritz	3200
Mary	3100
Tiago	3300

pd.Index

pd.Series

Deals with sequential, list-like data

Topics

- Creating Pandas Series and DataFrames
- Datatypes stored in Series and DataFrames
- Selection and Indexing in Series and DataFrames
- Operations on Series, incl. arithmetic and statistical measures
- Operations on DataFrames, incl. concatenation, joining, aggregation
- Importing and Exporting to other formats (.csv and Excel)

Packages required for this lesson

You will need **numpy** and **pandas** installed to your virtual environment (venv).

Windows:

```
.\.venv\Scripts\Activate

pip install numpy pandas
```

MacOS or Linux:

```
.venv/bin/activate

pip install numpy pandas
```

Activate your virtual environment within the Jupyter Notebook in VSCode.

Let’s go to our Jupyter Notebook…

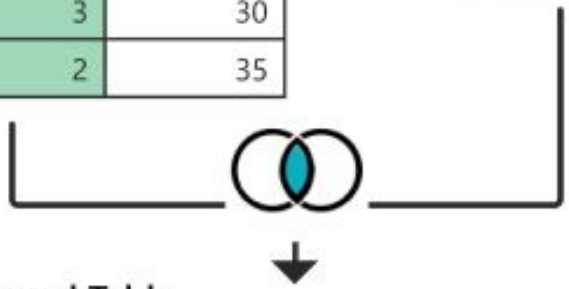
Table Inner Join Operation

Left Table

Date	CountryID	Units
1/1/2024	1	40
1/2/2024	1	25
1/3/2024	3	30
1/4/2024	2	35

Right Table

ID	Country
3	Panama
4	Spain



Merged Table

Date	CountryID	Units	Country
1/3/2024	3	30	Panama

Other Join Operations